

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 13, 2001, 14:14:35 ; Search time 229.28 Seconds

(without alignments)  
1.995 Million cell updates/sec

Title: PCT-US01-05825A-12

Perfect score: 42

Sequence: 1 GRLVQDC 8

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 390729 seqs, 57163235 residues

Total number of hits satisfying chosen parameters: 390729

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 08

Maximum Match 1008  
Listing first 45 summaries

Database :

A\_Geneseq\_0401.\*  
1: /SID6/gcgdata/geneseq/AA1980.DAT.\*  
2: /SID6/gcgdata/geneseq/AA1981.DAT.\*  
3: /SID6/gcgdata/geneseq/AA1982.DAT.\*  
4: /SID6/gcgdata/geneseq/AA1983.DAT.\*  
5: /SID6/gcgdata/geneseq/AA1984.DAT.\*  
6: /SID6/gcgdata/geneseq/AA1985.DAT.\*  
7: /SID6/gcgdata/geneseq/AA1986.DAT.\*  
8: /SID6/gcgdata/geneseq/AA1987.DAT.\*  
9: /SID6/gcgdata/geneseq/AA1988.DAT.\*  
10: /SID6/gcgdata/geneseq/AA1989.DAT.\*  
11: /SID6/gcgdata/geneseq/AA1990.DAT.\*  
12: /SID6/gcgdata/geneseq/AA1991.DAT.\*  
13: /SID6/gcgdata/geneseq/AA1992.DAT.\*  
14: /SID6/gcgdata/geneseq/AA1993.DAT.\*  
15: /SID6/gcgdata/geneseq/AA1994.DAT.\*  
16: /SID6/gcgdata/geneseq/AA1995.DAT.\*  
17: /SID6/gcgdata/geneseq/AA1996.DAT.\*  
18: /SID6/gcgdata/geneseq/AA1997.DAT.\*  
19: /SID6/gcgdata/geneseq/AA1998.DAT.\*  
20: /SID6/gcgdata/geneseq/AA1999.DAT.\*  
21: /SID6/gcgdata/geneseq/AA2000.DAT.\*  
22: /SID6/gcgdata/geneseq/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	42	100.0	8	21	Y79116
2	37	88.1	8	21	Y79114
3	35	83.3	8	21	Y79115
4	35	83.3	8	21	Y79128
5	35	83.3	246	14	R40924
6	34	81.0	146	21	G36361
7	34	81.0	156	21	G36360
8	34	81.0	176	21	G36359
9	33	78.6	8	21	Y79108
10	33	78.6	169	21	B25187
11	33	78.6	207	12	R13897

12	33	78.6	207	18	W14160	R. rhodochrous nit
13	33	78.6	207	20	Y02291	Alpha chain of a
14	32	76.2	8	21	Y79112	Peptide antagonist
15	32	76.2	77	21	B54158	Human pancreatic
16	32	76.2	238	19	W82389	Fllea saliva protei
17	32	76.2	242	19	W82392	Fllea saliva protei
18	32	76.2	264	19	W82390	Fllea saliva protei
19	31	73.8	117	18	W09067	Banana bunchy top
20	31	73.8	117	18	W09070	Banana bunchy top
21	31	73.8	553	20	Y35682	C. pneumoniae prot
22	31	73.8	963	17	R84082	Thermotabile enzym
23	31	73.8	1516	22	B11435	C. albicans sterol
24	31	73.8	1518	20	Y16099	Acetobacter xylinu
25	30	71.4	8	20	Y16868	Heat shock protein
26	30	71.4	8	21	Y79113	Peptide antagonist
27	30	71.4	8	21	Y79112	Peptide antagonist
28	30	71.4	116	16	R66349	Human immunoglobul
29	30	71.4	154	21	B32388	Human secreted pro
30	30	71.4	165	20	Y29528	Human lung tumour
31	30	71.4	165	21	B44423	Human lung tumour
32	30	71.4	240	22	B45991	Human MUC-1 scFv c
33	30	71.4	240	22	B45992	Human MUC-1 scFv c
34	30	71.4	240	22	B45993	Human MUC-1 scFv c
35	30	71.4	240	22	B45994	Human MUC-1 scFv c
36	30	71.4	240	22	B45995	Human MUC-1 scFv c
37	30	71.4	240	22	B45996	Human MUC-1 scFv c
38	30	71.4	240	22	B45997	Human MUC-1 scFv c
39	30	71.4	240	22	B45998	Human MUC-1 scFv c
40	30	71.4	240	22	B45999	Human MUC-1 scFv c
41	30	71.4	240	22	B46000	Human MUC-1 scFv c
42	30	71.4	240	22	B46001	Human MUC-1 scFv c
43	30	71.4	240	22	B46003	Human MUC-1 scFv c
44	30	71.4	240	22	B46004	Human MUC-1 scFv c
45	30	71.4	240	22	B46005	Human MUC-1 scFv c

#### ALIGNMENTS

RESULT	1
Y79116	Y79116 standard; Peptide: 8 AA.
XX	XX
AC	Y79116:
XX	XX
DT	05-JUN-2000 (first entry)
XX	XX
DE	Peptide antagonist of zonulin.
XX	XX
KW	Zonulin; antagonist; zonula occludens toxin receptor;
KW	blood-brain barrier; antiinflammatory; cerebroprotective;
KW	neuroprotective; dermatological; antitumor; antiviral;
KW	antibacterial; cytostatic; anti-HIV; vulnery; antiallergic;
KW	hypotensive; immunosuppressive; antiparasitic; vasotropic;
KW	gastrointestinal inflammation; therapy.
XX	XX
OS	Synthetic.
XX	XX
PN	WO200007609-A1.
XX	XX
PD	17-FEB-2000.
XX	XX
PF	28-JUL-1999; 99WO-US16683.
XX	XX
PR	03-AUG-1998; 98US-0127815.
XX	XX
PA	(UYMA-) UNIV MARYLAND BALTIMORE.
XX	XX
PI	Fasano A;
XX	XX
DR	WPI; 2000-205565/18.
XX	XX
PT	New peptide antagonist of zonulin useful as antiinflammatory agent for

PT treating cerebral ischemia, stroke, cerebral edema, gastritis,  
 PT shigellosis, viral gastroenteritis, meningitis, encephalomyelitis  
 PS Claim 1; Page 44; 69pp; English.

CC This present sequence is that of a peptide antagonist of zonulin  
 CC (2), one of 25 such peptides (see Y79105-29) of the invention,  
 CC which bind to a zonula occludens toxin (ZOT) receptor, yet do not  
 CC physiologically modulate the opening of mammalian tight junctions  
 CC (TJ). The peptide antagonists are based on a common motif of ZOT  
 CC binding. They can be prepared by chemical synthesis or by use of  
 CC recombinant DNA techniques. The peptide antagonists are used as an  
 CC antiinflammatory agents in the treatment of gastrointestinal  
 CC inflammation, where they bind to the ZOT receptor in the intestine  
 CC and yet does not physiologically modulate the opening of TJ in the  
 CC intestine. Gastrointestinal inflammation conditions give rise to  
 CC increased intestinal permeability and the peptide is useful for  
 CC treating intestinal conditions that cause protein losing enteropathy  
 CC caused by infection, e.g. Clostridium difficile infection,  
 CC enterocolitis, shigellosis, viral gastroenteritis, parasite  
 CC infestation, bacterial overgrowth, whipple's disease, diseases with  
 CC mucosal erosion or ulcerations, e.g. gastritis, gastric cancer,  
 CC collagenous colitis, inflammatory bowel disease, diseases marked by  
 CC lymphatic obstruction, e.g. congenital intestinal lymphangiectasia,  
 CC sarcoïdosis lymphoma, mesenteric tuberculosis, and after surgical  
 CC correction of congenital heart disease with Fontan's operation,  
 CC mucosal diseases without ulceration, e.g. Menetrier's disease,  
 CC coeliac disease, eosinophilic gastroenteritis, and immune diseases,  
 CC e.g. systemic lupus erythematosus or food allergies, primarily to  
 CC milk.

SO Sequence 8 AA;

Query Match 100.0%; Score 42; DB 21; Length 8;  
 Best Local Similarity 100.0%; Pred. No. 3.2e+05;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GRGLVODG 8  
 ||| |||||  
 DB 1 grglvqdg 8

RESULT 2  
 ID Y79114 standard; Peptide: 8 AA.  
 XX  
 AC Y79114;

DT 05-JUN-2000 (first entry)  
 XX

DE Peptide antagonist of zonulin.  
 XX

KW Zonulin; antagonist; zonula occludens toxin receptor;  
 KW blood-brain barrier; antiinflammatory; cerebroprotective;  
 KW neuroprotective; dermatological; antiulcer; antiviral;  
 KW antibacterial; cytostatic; anti-HIV; vulnary; antiallergic;  
 KW hypotensive; immunosuppressive; antiparasitic; vasotropic;  
 KW gastrointestinal inflammation; therapy.  
 XX

OS Synthetic.  
 XX

PN WO200007609-A1.  
 XX

PD 17-FEB-2000.  
 XX

PF 28-JUL-1999; 99WO-US16683.  
 XX

PR 03-AUG-1998; 98US-0127815.  
 XX

PA (UTMA-) UNIV MARYLAND BALTIMORE.  
 XX

PI Fasano A;  
 DR WPI: 2000-205565/18.  
 XX

PT New peptide antagonist of zonulin useful as antiinflammatory agent for  
 PT treating cerebral ischemia, stroke, cerebral edema, gastritis,  
 PT shigellosis, viral gastroenteritis, meningitis, encephalomyelitis  
 PS Claim 1; Page 44; 69pp; English.

CC This present sequence is that of a peptide antagonist of zonulin  
 CC (2), one of 25 such peptides (see Y79105-29) of the invention,  
 CC which bind to a zonula occludens toxin (ZOT) receptor, yet do not  
 CC physiologically modulate the opening of mammalian tight junctions  
 CC (TJ). The peptide antagonists are based on a common motif of ZOT  
 CC binding. They can be prepared by chemical synthesis or by use of  
 CC recombinant DNA techniques. The peptide antagonists are used as an  
 CC antiinflammatory agents in the treatment of gastrointestinal  
 CC inflammation, where they bind to the ZOT receptor in the intestine  
 CC and yet does not physiologically modulate the opening of TJ in the  
 CC intestine. Gastrointestinal inflammation conditions give rise to  
 CC increased intestinal permeability and the peptide is useful for  
 CC treating intestinal conditions that cause protein losing enteropathy  
 CC caused by infection, e.g. Clostridium difficile infection,  
 CC enterocolitis, shigellosis, viral gastroenteritis, parasite  
 CC infestation, bacterial overgrowth, whipple's disease, diseases with  
 CC mucosal erosion or ulcerations, e.g. gastritis, gastric cancer,  
 CC collagenous colitis, inflammatory bowel disease, diseases marked by  
 CC lymphatic obstruction, e.g. congenital intestinal lymphangiectasia,  
 CC sarcoïdosis lymphoma, mesenteric tuberculosis, and after surgical  
 CC correction of congenital heart disease with Fontan's operation,  
 CC mucosal diseases without ulceration, e.g. Menetrier's disease,  
 CC coeliac disease, eosinophilic gastroenteritis, and immune diseases,  
 CC e.g. systemic lupus erythematosus or food allergies, primarily to  
 CC milk.

SO Sequence 8 AA;

Query Match 88.1%; Score 37; DB 21; Length 8;  
 Best Local Similarity 87.5%; Pred. No. 3.2e+05;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GRGLVODG 8  
 ||| |||||  
 DB 1 grgcvgdg 8

RESULT 3  
 ID Y79115 standard; Peptide: 8 AA.  
 XX  
 AC Y79115;

DT 05-JUN-2000 (first entry)  
 XX

DE Peptide antagonist of zonulin.  
 XX

KW Zonulin; antagonist; zonula occludens toxin receptor;  
 KW blood-brain barrier; antiinflammatory; cerebroprotective;  
 KW neuroprotective; dermatological; antiulcer; antiviral;  
 KW antibacterial; cytostatic; anti-HIV; vulnary; antiallergic;  
 KW hypotensive; immunosuppressive; antiparasitic; vasotropic;  
 KW gastrointestinal inflammation; therapy.  
 XX

OS Synthetic.  
 XX

PN WO200007609-A1.  
 XX

PD 17-FEB-2000.  
 XX

PF 28-JUL-1999; 99WO-US16683.  
 XX



```

XX 03-AUG-1998: 98US-0127815.
XX (UYMA-) UNIV MARYLAND BALTIMORE.
XX Fasano A:
XX WPI: 2000-205565/18.
XX
XX New peptide antagonist of zonulin useful as antiinflammatory agent for
XX treating cerebral ischemia, stroke, cerebral edema, gastritis,
XX shigellosis, viral gastroenteritis, meningitis, encephalomyelitis
XX
XX Claim 1: Page 44: 69pp; English.
XX
XX This present sequence is that of a peptide antagonist of zonulin
XX (Z), one of 25 such peptides (see Y79105-29) of the invention,
XX which bind to a zonula occludens toxin (ZOT) receptor, yet do not
XX physiologically modulate the opening of mammalian tight junctions
XX (TJ). The peptide antagonists are based on a common motif of ZOT
XX and human zonulins, which is believed to be critical for receptor
XX binding. They can be prepared by chemical synthesis or by use of
XX recombinant DNA techniques. The peptide antagonists are used as an
XX antiinflammatory agents in the treatment of gastrointestinal
XX inflammation, where they bind to the ZOT receptor in the intestine
XX and yet does not physiologically modulate the opening of TJ in the
XX intestine. Gastrointestinal inflammation conditions give rise to
XX increased intestinal permeability and the peptide is useful for
XX treating intestinal conditions that cause protein losing enteropathy
XX caused by infection, e.g. Clostridium difficile infection,
XX enterocolitis, shigellosis, viral gastroenteritis, parasite
XX infection, bacterial overgrowth, whipple's disease, diseases with
XX mucosal erosion or ulcerations, e.g. gastritis, gastric cancer,
XX collagenous colitis, inflammatory bowel disease, diseases marked by
XX lymphatic obstruction, e.g. congenital intestinal lymphangiectasia,
XX sarcoidosis lymphoma, mesenteric tuberculosis, and after surgical
XX correction of congenital heart disease with Fontan's operation,
XX mucosal diseases without ulceration, e.g. Menetrier's disease,
XX coeliac disease, eosinophilic gastroenteritis, and immune diseases,
XX e.g. systemic lupus erythematosus or food allergies, primarily to
XX milk.
XX
XX Sequence 8 AA:
XX
XX Query Match 83.3%; Score 35; DB 21; Length 8;
XX Best Local Similarity 87.5%; Pred. No. 3.2e+05;
XX Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
XX 1 GRGLVODG 8
XX 1 grglvqpg 8
XX
XX RESULT 4
XX Y79128
XX ID Y79128 standard; Peptide: 8 AA.
XX AC Y79128;
XX DT 05-JUN-2000 (first entry)
XX DE Peptide antagonist of zonulin.
XX
XX Zonulin: antagonist; zonula occludens toxin receptor;
XX blood-brain barrier; antiinflammatory; cerebroprotective;
XX neuroprotective; dermatological; antiulcer; antiviral;
XX antibacterial; cytostatic; anti-HIV; vulnerability; antiallergic;
XX hypotensive; immunosuppressive; antiparasitic; vasotropic;
XX gastrointestinal inflammation; therapy.
XX
XX Synthetic.
XX

```

```

PN WO200007609-A1.
XX
XX 17-FEB-2000.
XX
XX 28-JUL-1999: 99WO-US16683.
XX
XX 03-AUG-1998: 98US-0127815.
XX
XX (UYMA-) UNIV MARYLAND BALTIMORE.
XX Fasano A:
XX WPI: 2000-205565/18.
XX
XX New peptide antagonist of zonulin useful as antiinflammatory agent for
XX treating cerebral ischemia, stroke, cerebral edema, gastritis,
XX shigellosis, viral gastroenteritis, meningitis, encephalomyelitis
XX
XX Claim 1: Page 48: 69pp; English.
XX
XX This present sequence is that of a peptide antagonist of zonulin
XX (Z), one of 25 such peptides (see Y79105-29) of the invention,
XX which bind to a zonula occludens toxin (ZOT) receptor, yet do not
XX physiologically modulate the opening of mammalian tight junctions
XX (TJ). The peptide antagonists are based on a common motif of ZOT
XX and human zonulins, which is believed to be critical for receptor
XX binding. They can be prepared by chemical synthesis or by use of
XX recombinant DNA techniques. The peptide antagonists are used as an
XX antiinflammatory agents in the treatment of gastrointestinal
XX inflammation, where they bind to the ZOT receptor in the intestine
XX and yet does not physiologically modulate the opening of TJ in the
XX intestine. Gastrointestinal inflammation conditions give rise to
XX increased intestinal permeability and the peptide is useful for
XX treating intestinal conditions that cause protein losing enteropathy
XX caused by infection, e.g. Clostridium difficile infection,
XX enterocolitis, shigellosis, viral gastroenteritis, parasite
XX infection, bacterial overgrowth, whipple's disease, diseases with
XX mucosal erosion or ulcerations, e.g. gastritis, gastric cancer,
XX collagenous colitis, inflammatory bowel disease, diseases marked by
XX lymphatic obstruction, e.g. congenital intestinal lymphangiectasia,
XX sarcoidosis lymphoma, mesenteric tuberculosis, and after surgical
XX correction of congenital heart disease with Fontan's operation,
XX mucosal diseases without ulceration, e.g. Menetrier's disease,
XX coeliac disease, eosinophilic gastroenteritis, and immune diseases,
XX e.g. systemic lupus erythematosus or food allergies, primarily to
XX milk.
XX
XX Sequence 8 AA:
XX
XX Query Match 83.3%; Score 35; DB 21; Length 8;
XX Best Local Similarity 87.5%; Pred. No. 3.2e+05;
XX Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
XX 1 GRGLVODG 8
XX 1 grglvqpg 8
XX
XX RESULT 5
XX R40924
XX ID R40924 standard; Protein: 246 AA.
XX AC R40924;
XX DT 17-FEB-1994 (first entry)
XX DE Protein able to bind to HIV-1 tat protein.
XX
XX Antibodies; antigen binding proteins; library; HIV;
XX Human Immunodeficiency Virus.
XX
XX Synthetic.
XX

```



```
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144332.
PR 20-JUL-1999; 99US-0144332.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 23-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 26-JUL-1999; 99US-0145224.
PR 27-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 06-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149358.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 26-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 13-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157717.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.

PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 81.0%; Score 34; DB 21; Length 146;
Best Local Similarity 62.5%; Pred. No. 31;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 1 GRGLVQDG 8
Db 25 gkylirdg 32

RESULT 7
ID G36360 standard; Protein: 166 AA.
XX
AC G36360;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 44548.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
hybridisation assay; genetic mapping; gene expression control; promoter;
termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
PF
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
```

PR	28-APR-1999;	990S-0131449.
PR	30-APR-1999;	990S-0132048.
PR	30-APR-1999;	990S-0132407.
PR	04-MAY-1999;	990S-0132484.
PR	05-MAY-1999;	990S-0132485.
PR	06-MAY-1999;	990S-0132486.
PR	06-MAY-1999;	990S-0132487.
PR	07-MAY-1999;	990S-0132487.
PR	11-MAY-1999;	990S-0134256.
PR	14-MAY-1999;	990S-0134218.
PR	14-MAY-1999;	990S-0134221.
PR	14-MAY-1999;	990S-0134221.
PR	18-MAY-1999;	990S-0134370.
PR	18-MAY-1999;	990S-0134376.
PR	19-MAY-1999;	990S-0134378.
PR	20-MAY-1999;	990S-0134911.
PR	21-MAY-1999;	990S-0135314.
PR	24-MAY-1999;	990S-0135333.
PR	25-MAY-1999;	990S-0135629.
PR	25-MAY-1999;	990S-0136021.
PR	27-MAY-1999;	990S-0136392.
PR	28-MAY-1999;	990S-0136782.
PR	01-JUN-1999;	990S-0137222.
PR	03-JUN-1999;	990S-0137528.
PR	04-JUN-1999;	990S-0137602.
PR	07-JUN-1999;	990S-0137724.
PR	08-JUN-1999;	990S-0137724.
PR	10-JUN-1999;	990S-0138094.
PR	14-JUN-1999;	990S-0138540.
PR	14-JUN-1999;	990S-0138847.
PR	16-JUN-1999;	990S-0139119.
PR	16-JUN-1999;	990S-0139452.
PR	17-JUN-1999;	990S-0139455.
PR	18-JUN-1999;	990S-0139459.
PR	18-JUN-1999;	990S-0139459.
PR	18-JUN-1999;	990S-0139459.
PR	18-JUN-1999;	990S-0139456.
PR	18-JUN-1999;	990S-0139457.
PR	18-JUN-1999;	990S-0139458.
PR	18-JUN-1999;	990S-0139459.
PR	18-JUN-1999;	990S-0139460.
PR	18-JUN-1999;	990S-0139461.
PR	18-JUN-1999;	990S-0139462.
PR	18-JUN-1999;	990S-0139463.
PR	18-JUN-1999;	990S-0139750.
PR	21-JUN-1999;	990S-0139763.
PR	22-JUN-1999;	990S-0139817.
PR	23-JUN-1999;	990S-0139859.
PR	23-JUN-1999;	990S-0140353.
PR	24-JUN-1999;	990S-0140354.
PR	28-JUN-1999;	990S-0140695.
PR	29-JUN-1999;	990S-0140823.
PR	30-JUN-1999;	990S-0140991.
PR	01-JUL-1999;	990S-0141287.
PR	01-JUL-1999;	990S-0141842.
PR	02-JUL-1999;	990S-0142151.
PR	06-JUL-1999;	990S-0142055.
PR	09-JUL-1999;	990S-0142390.
PR	09-JUL-1999;	990S-0142803.
PR	12-JUL-1999;	990S-0142920.
PR	13-JUL-1999;	990S-0143542.
PR	14-JUL-1999;	990S-0143624.
PR	15-JUL-1999;	990S-0144005.
PR	16-JUL-1999;	990S-0144085.
PR	19-JUL-1999;	990S-0144086.
PR	19-JUL-1999;	990S-0144331.
PR	19-JUL-1999;	990S-0144332.
PR	19-JUL-1999;	990S-0144333.
PR	19-JUL-1999;	990S-0144334.
PR	19-JUL-1999;	990S-0144335.
PR	20-JUL-1999;	990S-0144352.
PR	20-JUL-1999;	990S-0144632.
PR	20-JUL-1999;	990S-0144884.
PR	21-JUL-1999;	990S-0144814.
PR	21-JUL-1999;	990S-0145086.
PR	22-JUL-1999;	990S-0145088.
PR	22-JUL-1999;	990S-0145089.
PR	22-JUL-1999;	990S-0145089.
PR	22-JUL-1999;	990S-0145119.
PR	22-JUL-1999;	990S-0145119.
PR	23-JUL-1999;	990S-0145119.
PR	23-JUL-1999;	990S-0145224.
PR	26-JUL-1999;	990S-0145276.
PR	27-JUL-1999;	990S-0145313.
PR	27-JUL-1999;	990S-0145313.
PR	28-JUL-1999;	990S-0145313.
PR	02-AUG-1999;	990S-0145313.
PR	02-AUG-1999;	990S-0145313.
PR	03-AUG-1999;	990S-0145313.
PR	04-AUG-1999;	990S-0145313.
PR	04-AUG-1999;	990S-0145313.
PR	05-AUG-1999;	990S-0145313.
PR	05-AUG-1999;	990S-0145313.
PR	06-AUG-1999;	990S-0145313.
PR	06-AUG-1999;	990S-0145313.
PR	09-AUG-1999;	990S-0145313.
PR	09-AUG-1999;	990S-0145313.
PR	10-AUG-1999;	990S-0145313.
PR	11-AUG-1999;	990S-0145313.
PR	12-AUG-1999;	990S-014

PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 81.0%; Score 34; DB 21; Length 166;  
Best Local Similarity 62.5%; Pred. No. 35;  
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 1 GRGLVODG 8  
|:|:|:|:|  
Db 45 gkxgltirdg 52

RESULT 8  
G36359  
ID G36359 standard; Protein; 176 AA.  
XX AC G36359;  
XX 18-OCT-2000 (first entry)  
DT Arabidopsis thaliana protein fragment SEQ ID NO: 44547.  
XX  
XX Protein identification; signal transduction pathway; metabolic pathway;  
KM hybridisation assay; genetic mapping; gene expression control; promoter;  
KM termination sequence.  
XX  
OS Arabidopsis thaliana.  
XX  
PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
XX 25-FEB-2000; 2000EP-0301439.  
PE  
XX 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.

PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 01-JUN-1999; 99US-0137528.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 17-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 28-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144884.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.

PR	23-JUL-1999;	99US-0145145
PR	23-JUL-1999;	99US-0145218
PR	23-JUL-1999;	99US-0145224
PR	26-JUL-1999;	99US-0145276
PR	27-JUL-1999;	99US-0145313
PR	27-JUL-1999;	99US-0145918
PR	28-JUL-1999;	99US-0145951
PR	02-AUG-1999;	99US-0146388
PR	02-AUG-1999;	99US-0146389
PR	03-AUG-1999;	99US-0147038
PR	04-AUG-1999;	99US-0147204
PR	05-AUG-1999;	99US-0147302
PR	05-AUG-1999;	99US-0147292
PR	06-AUG-1999;	99US-0147260
PR	06-AUG-1999;	99US-0147303
PR	09-AUG-1999;	99US-0147436
PR	09-AUG-1999;	99US-0147483
PR	10-AUG-1999;	99US-0147935
PR	11-AUG-1999;	99US-0148171
PR	12-AUG-1999;	99US-0148319
PR	13-AUG-1999;	99US-0148341
PR	13-AUG-1999;	99US-0148684
PR	16-AUG-1999;	99US-0149368
PR	17-AUG-1999;	99US-0149175
PR	18-AUG-1999;	99US-0149426
PR	20-AUG-1999;	99US-0149722
PR	20-AUG-1999;	99US-0149723
PR	23-AUG-1999;	99US-0149929
PR	23-AUG-1999;	99US-0149902
PR	25-AUG-1999;	99US-0149930
PR	26-AUG-1999;	99US-0150566
PR	27-AUG-1999;	99US-0150884
PR	27-AUG-1999;	99US-0151065
PR	27-AUG-1999;	99US-0151066
PR	30-AUG-1999;	99US-0151080
PR	31-AUG-1999;	99US-0151303
PR	01-SEP-1999;	99US-0151348
PR	07-SEP-1999;	99US-0151930
PR	10-SEP-1999;	99US-0152363
PR	13-SEP-1999;	99US-0153070
PR	15-SEP-1999;	99US-0153758
PR	16-SEP-1999;	99US-0154039
PR	20-SEP-1999;	99US-0154039
PR	22-SEP-1999;	99US-0155139
PR	22-SEP-1999;	99US-0155139
PR	22-SEP-1999;	99US-0155486
PR	28-SEP-1999;	99US-0156458
PR	29-SEP-1999;	99US-0156596
PR	04-OCT-1999;	99US-0157117
PR	05-OCT-1999;	99US-0157153
PR	06-OCT-1999;	99US-0157865
PR	07-OCT-1999;	99US-0158029
PR	08-OCT-1999;	99US-0158029
PR	12-OCT-1999;	99US-0158332
PR	13-OCT-1999;	99US-0158332
PR	13-OCT-1999;	99US-0158293
PR	14-OCT-1999;	99US-0158294
PR	14-OCT-1999;	99US-0159259
PR	14-OCT-1999;	99US-0159330
PR	14-OCT-1999;	99US-0159331
PR	14-OCT-1999;	99US-0159637
PR	14-OCT-1999;	99US-0159637
PR	18-OCT-1999;	99US-0159584
PR	21-OCT-1999;	99US-0160741
PR	21-OCT-1999;	99US-0160767
PR	21-OCT-1999;	99US-0160768
PR	21-OCT-1999;	99US-0160770
PR	21-OCT-1999;	99US-0160814
PR	21-OCT-1999;	99US-0160815

PR 22-OCT-1999; 99US-0160980  
PR 22-OCT-1999; 99US-0160981  
PR 22-OCT-1999; 99US-0160989  
PR 25-OCT-1999; 99US-0161404  
PR 25-OCT-1999; 99US-0161405  
PR 25-OCT-1999; 99US-0161406  
PR 26-OCT-1999; 99US-0161359  
PR 26-OCT-1999; 99US-0161360  
PR 26-OCT-1999; 99US-0161361  
PR 28-OCT-1999; 99US-0161920  
PR 28-OCT-1999; 99US-0161992  
PR 28-OCT-1999; 99US-0161993  
PR 29-OCT-1999; 99US-0162142.

Query Match	81.0%;	Score 34;	DB 21;	Length 176;
Best Local Similarity	62.5%;	Pred. No. 37;		
Matches	5;	Conservative 3;	Mismatches 0;	Indels 0;
Qy	1 GRLVDG 8			
	: : : :			
Db	55 gkgldtg 62			

RESULT	9
Y79108	.
ID	Y79108 standard; Peptide; 8 AA.
XX	

05-JUN-2000 (first entry)  
Peptide antagonist of zonulin

gastrointestinal inflammation; therapy.

XX	synthetic.	
DS		
XX	WO200007609-A1.	
NN		
XX	17-FEB-2000.	
DD		
XX		
FF	28-JUL-1999;	99WO-US16683.
XX		
RR	03-AUG-1998;	98US-0127815.
XX		
XX	(UYNA-) UNIV MARYLAND BALTIMORE	

WPI; 2000-205565/18.

new peptide antagonist of zonulin useful as antiinflammatory agent for treating cerebral ischemia, stroke, cerebral edema, gastritis, shigellosis, viral gastroenteritis, meningitis, encephalomyelitis - Claim 1; Page 42; 69pp; English.

(2), one of 25 such peptides, or a peptide antagonist of zonulin which bind to a zonula occludens toxin (ZOT) receptor, yet do not physiologically modulate the opening of mammalian tight junctions (TJ). The peptide antagonists are based on a common motif of ZOT and human zonulins, which is believed to be critical for receptor binding. They can be prepared by chemical synthesis or by use of recombinant DNA techniques. The peptide antagonists are used as anti-inflammatory agents in the treatment of gastrointestinal inflammation, where they bind to the ZOT receptor in the intestine and yet does not physiologically modulate the opening of TJ in the

CC intestine. Gastrointestinal inflammation conditions give rise to  
CC increased intestinal permeability and the peptide is useful for  
CC treating intestinal conditions that cause protein losing enteropathy  
CC caused by infection, e.g. Clostridium difficile infection,  
CC enterocolitis, shigellosis, viral gastroenteritis, parasite  
CC infection, bacterial overgrowth, whipple's disease, diseases with  
CC mucosal erosion or ulcerations, e.g. gastritis, gastric cancer,  
CC collagenous colitis, inflammatory bowel disease, diseases marked by  
CC lymphatic obstruction, e.g. congenital intestinal lymphangiectasia,  
CC sarcoidosis lymphoma, mesenteric tuberculosis, and after surgical  
CC correction of congenital heart disease with Fontan's operation,  
CC mucosal diseases without ulceration, e.g. Menetrier's disease,  
CC coeliac disease, eosinophilic gastroenteritis, and immune diseases,  
CC e.g. systemic lupus erythematosus or food allergies, primarily to  
CC milk.  
CC  
CC  
SQ Sequence 8 AA;  
  
Query Match 78.6%; Score 33; DB 21; Length 8;  
Best Local Similarity 87.5%; Pred. No. 3.2e+05;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
  
OY 1 GRLVODG 8  
1111111  
Db 1 grlvqdg 8  
  
RESULT 10  
B25187  
ID B25187 standard; Protein: 169 AA.  
AC B25187;  
XX  
XX 27-NOV-2000 (first entry)  
DT  
XX  
XX Eucalyptus grandis cell signalling involved protein SEQ ID NO:506.  
DE  
XX  
XX Eucalyptus grandis; Pinus radiata; Monterey pine; plant; modification;  
KW plant cell signalling; modulation; transgenic plant; pathogen; growth;  
KW environmental change; development; cell proliferation; differentiation;  
KW elongation; survival; disease resistance; nutrient metabolism.  
XX  
XX Eucalyptus grandis.  
OS  
XX  
XX WO200042171-A1.  
PN  
XX  
XX 20-JUL-2000.  
PD  
XX  
XX 11-JAN-2000; 2000MO-US00724.  
PF  
XX  
XX 12-JAN-1999; 99US-0228986.  
PR  
XX  
XX 01-NOV-1999; 99US-0162866.  
PR  
XX  
XX (GENE-) GENESIS RES & DEV CORP LTD.  
PA  
XX  
XX Strabala TJ, Nieuwenhuizen NJ;  
PI  
XX  
XX WPI; 2000-476052/41.  
DR  
XX  
XX Isolated polynucleotide encoding a polypeptide involved in cell  
PT signalling used for generating transgenic plants with modified responses  
PT to external signals -  
PT  
XX  
XX Claim 3; Page 224; 527pp; English.  
PS  
XX  
XX A79263 to A79736 and B25100 to B25570 represent polynucleotide and  
CC protein sequences isolated from eucalyptus (Eucalyptus grandis) or pine  
CC (Pinus radiata also known as Monterey pine). The protein sequences are  
CC involved in cell signalling. The polynucleotide and protein sequences  
CC can be used to modify the response of plant cells to external signals  
CC e.g. environmental changes or pathogens during the growth and development  
CC of a plant. They can be used to modify cell proliferation,

CC differentiation, elongation and survival, resistance to disease and  
CC nutrient metabolism. Examples of modifications which can be produced are  
CC altered fruit ripening and senescence of leaves and flowers e.g. to  
CC delay senescence and prolong the life of cut flowers or enhance  
CC senescence of reproductive organs to engineer sterile plants. Other  
CC modifications of reproductive organs to engineer sterile cell types or  
CC organs providing fruit and vegetables which have a longer shelf life  
CC between harvest and consumption, or to decrease branching frequency in  
CC forest tree species giving long stretches of valuable knot-free clear  
CC wood which can be used in solid timber furniture and veneers.  
CC  
CC  
SQ Sequence 169 AA;  
  
Query Match 78.6%; Score 33; DB 21; Length 169;  
Best Local Similarity 85.7%; Pred. No. 55;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
  
OY 2 RGLVODG 8  
1111111  
Db 46 rgllvqdg 52  
  
RESULT 11  
R13897  
ID R13897 standard; Protein: 207 AA.  
XX  
XX R13897;  
AC  
XX  
XX 25-NOV-1991 (first entry)  
DT  
XX  
XX Nitrile hydratase alpha subunit.  
DE  
XX  
XX NH; probe; subunit; alpha; beta.  
KW  
XX  
XX Rhodococcus sp. N-774.  
OS  
XX  
XX EP445646-A.  
PN  
XX  
XX 11-SEP-1991.  
PD  
XX  
XX 27-FEB-1991; 91EP-0102937.  
PF  
XX  
XX 28-FEB-1990; 90JP-0048078.  
PR  
XX  
XX (NITP ) NITTO CHEM IND KR.  
PA  
XX  
XX Teruhiko B, Hideaki Y, Toru N, Sueharu H, Makoto N;  
PI  
XX  
XX WPI; 1991-268533/37.  
DR  
XX  
XX N-PSDB; Q13618.  
DR  
XX  
XX DNA encoding two types of nitrile hydratase - from Rhodococcus  
PT rhodochrous J-I, used to produce recombinant enzyme for hydration  
PT of e.g. benzonitrile  
PT  
XX  
XX Disclosure; Page 20-23; 30pp; English.  
PS  
XX  
XX A DNA probe is prepd. from JM105/pYUK121 (FERM BP-1937) described in  
CC JP-A-2-119778/1990 due to the high degree of homology in the amino acid  
CC sequence between the nitrile hydratase beta subunit of Rhodococcus  
CC sp. N-774 and Rhodococcus rhodochrous J-I.  
CC Plasmid pYUK121 contg. nitrile hydratase gene derived from Rhodococcus  
CC sp. N-774 is prepd. from a JM105/pYUK121 culture. pYUK121 DNA is  
CC digested with SphI and SalI. This fragment contains the nitrile  
CC hydratase gene which is radiolabelled for use as probe.  
CC The NH gene encodes an alpha subunit as represented here, and a  
CC beta subunit represented in R13898.  
CC See also Q13618-20.  
CC  
XX  
XX Sequence 207 AA;

Query Match 78.6%; Score 33; DB 12; Length 207;  
Best Local Similarity 75.0%; Pred. No. 68;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GRGLVODG 8  
1:|||||  
Db 32 gkglyvpdg 39

## RESULT 12

ID W14160 standard; Protein: 207 AA.  
AC W14160;

DT 24-JUN-1997 (first entry)  
XX

DE R. rhodochrous nitrile hydratase alpha subunit.  
XX

KW Rhodococcus rhodochrous; amidase; probe: amplification; primer: PCR;  
KW polymerase chain reaction; phage; open reading frame; nitrile hydratase;  
XX enzyme; genetic engineering.

OS Rhodococcus rhodochrous.  
XX

FI Key Location/Qualifiers  
FT Misc-difference 148 /note= "encoded by codon ATC"  
PN JP0900973-A.

PD 14-JAN-1997.  
XX

PF 27-JUN-1995; 95JP-0184934.  
XX

PR 27-JUN-1995; 95JP-0184934.  
XX

PA (CHCC) CHISSO CORP.  
XX

DR WPI: 1997-126430/12.  
XX

DR N-PSDB: T62386.  
XX

PT Rhodococcus rhodochrous nitrile hydratase and amidase genes  
XX useful for the industrial preparation of useful cpds.  
PS Claim 2; Page 7-8; 16pp; Japanese.

CC This is the amino acid sequence of the nitrile hydratase alpha subunit  
CC from Rhodococcus rhodochrous strain IF015564. The coding sequences were  
CC isolated by screening a genomic R. rhodochrous DNA library in EMBL3  
CC vector, using a probe amplified by the primers T62388-9. The primers  
CC amplified a 0.7 kb fragment of the amidase gene. The library screen  
CC isolated a phage containing a 15 kb insert. When analysed, the insert  
CC was seen to contain an approx. 6 kb SacI-EcoRI fragment (see T62387)  
CC containing the open reading frames for the amidase gene (T62385), the  
CC nitrile hydratase alpha and beta subunits and another open reading frame  
CC encoding a 399 amino acid protein (W14162) of unknown function.  
CC The enzymes, prepared by genetic engineering methods, can be used in the  
CC industrial production of useful compounds.  
XX  
SQ Sequence 207 AA;

Query Match 78.6%; Score 33; DB 18; Length 207;  
Best Local Similarity 75.0%; Pred. No. 68;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GRGLVODG 8  
1:|||||  
Db 32 gkglyvpdg 39

RESULT 13

Y02291  
ID Y02291 standard; protein: 207 AA.  
XX  
AC Y02291;  
XX

DT 14-JUL-1999 (first entry)  
XX

DE Alpha chain of a Brevibacterium nitrile hydratase.  
XX

KW Halogenated benzamide; enzymatic hydrolysis; halogenated benzonitrile;  
KW Brevibacterium; nitrile hydratase; tetrameric enzyme; alpha chain;  
KW beta chain; strain R312; mutant A4; 2,6-difluorobenzamide; intermediate;  
XX insecticide; diflubenzuron.

OS Brevibacterium sp.  
XX

PN FR2770214-A1.  
XX

PD 30-APR-1999.  
XX

PF 24-OCT-1997; 97FR-0013368.  
XX

PR 24-OCT-1997; 97FR-0013368.  
XX

PA (RHOD) RHODIA CHIM.  
XX

PI Chasseau L, Jourdat C, Petre D;  
XX

DR WPI: 1999-280394/24.  
XX

PT Halogenated benzamide production - by halogenated benzonitrile  
XX hydrolysis with Brevibacterium nitrile hydratase enzyme  
PS Claim 1; Page 9; 12pp; French.

CC The specification describes the production of halogenated benzamides  
CC by enzymatic hydrolysis of halogenated benzonitriles using a  
CC Brevibacterium nitrile hydratase. This enzyme is a tetrameric enzyme  
CC comprising 2 alpha and 2 beta chains. The present sequence represents  
CC the alpha chain. The enzyme and the Brevibacterium strain from which it  
CC is derived (strain R312 or its mutant A4) have higher activity for  
CC hydrolysing halogenated benzonitriles than for hydrolysing benzonitrile.  
CC The method is especially for preparing 2,6-difluorobenzamide, which is  
CC an intermediate for insecticides, especially diflubenzuron.  
XX  
SQ Sequence 207 AA;

Query Match 78.6%; Score 33; DB 20; Length 207;  
Best Local Similarity 75.0%; Pred. No. 68;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GRGLVODG 8  
1:|||||  
Db 32 gkglyvpdg 39

## RESULT 14

ID Y79112 standard; Peptide: 8 AA.  
XX

AC Y79112;  
XX

DT 05-JUN-2000 (first entry)  
XX

DE Peptide antagonist of zonulin.  
XX

KW Zonulin; antagonist; zonula occludens toxin receptor;  
KW blood-brain barrier; antiinflammatory; cerebroprotective;  
KW neuroprotective; dermatological; anticancer; antiviral;  
KW antibacterial; cytostatic; anti-HIV; vulnerrary; antiallergic;  
KW hypotensive; immunosuppressive; antiparasitic; vasotropic;  
KW gastrointestinal inflammation; therapy.



```

XX OS Synthetic.
XX PN WO200007609-A1.
XX PD 17-FEB-2000.
XX PF 28-JUL-1999; 99WO-US16683.
XX PR 03-AUG-1998; 98US-0127815.
XX PA (UYMA-) UNIV MARYLAND BALTIMORE.
XX PI Fasano A;
XX DR WPI; 2000-205565/18.
XX PT New peptide antagonist of zonulin useful as antiinflammatory agent for
XX PT treating cerebral ischemia, stroke, cerebral edema, gastritis,
XX PT shigellosis, viral gastroenteritis, meningitis, encephalomyelitis -
XX PS Claim 1; Page 43; 69pp; English.
XX CC This present sequence is that of a peptide antagonist of zonulin
XX CC (Z), one of 25 such peptides (see Y79105-29) of the invention,
XX CC which bind to a zonula occludens toxin (ZOT) receptor, yet do not
XX CC physiologically modulate the opening of mammalian tight junctions
XX CC (TJ). The peptide antagonists are based on a common motif of ZOT
XX CC and human zonulins, which is believed to be critical for receptor
XX CC binding. They can be prepared by chemical synthesis or by use of
XX CC recombinant DNA techniques. The peptide antagonists are used as an
XX CC antiinflammatory agents in the treatment of gastrointestinal
XX CC inflammation, where they bind to the ZOT receptor in the intestine
XX CC and yet does not physiologically modulate the opening of TJ in the
XX CC intestine. Gastrointestinal inflammation conditions give rise to
XX CC increased intestinal permeability and the peptide is useful for
XX CC treating intestinal conditions that cause protein losing enteropathy
XX CC caused by infection, e.g. Clostridium difficile infection,
XX CC enterocolitis, shigellosis, viral gastroenteritis, parasite
XX CC infestation, bacterial overgrowth, whipple's disease, diseases with
XX CC mucosal erosion or ulcerations, e.g. gastritis, gastric cancer,
XX CC collagenous colitis, inflammatory bowel disease, diseases marked by
XX CC lymphatic obstruction, e.g. congenital intestinal lymphangiectasia,
XX CC sarcoidosis lymphoma, mesenteric tuberculosis, and after surgical
XX CC correction of congenital heart disease with Fontan's operation,
XX CC mucosal diseases without ulceration, e.g. Menetrier's disease,
XX CC coeliac disease, eosinophilic gastroenteritis, and immune diseases,
XX CC e.g. systemic lupus erythematosus or food allergies, primarily to
XX CC milk.
XX SO Sequence 8 AA;

Query Match 76.2%; Score 32; DB 21; Length 8;
Best Local Similarity 87.5%; Pred. No. 3.2e+05;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GRLGVODG 8
   1111111
Db 1 grllvqdg 8

RESULT 15
B54158 B54158 standard; Protein; 77 AA.
XX AC B54158;
XX DT 09-MAR-2001 (first entry)
XX DE Human pancreatic cancer antigen protein sequence SEQ ID NO:610.
XX KW Human; pancreas; pancreatic cancer; pancreatic cancer antigen;

```

```

KW detection; diagnosis; identification; cytostatic; neuroprotective;
KW nootropic; immunomodulatory; relaxant; contraceptive; gynaecological;
KW antiinflammatory; cardiant; gene therapy; chromosome mapping;
KW linkage analysis; tissue identification; tissue typing; forensic;
KW neural; immune system; muscular; reproductive; gastrointestinal;
KW pulmonary; cardiovascular; renal; proliferative.
XX OS Homo sapiens.
XX PN WO200055320-A1.
XX PD 21-SEP-2000.
XX PF 08-MAR-2000; 2000WO-US05989.
XX PR 12-MAR-1999; 99US-0124270.
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX PI Rosen CA, Ruben SM;
XX DR WPI; 2000-579444/54.
XX DR N-PSDB; C98923.
XX PT New nucleic acid that is a pancreatic cancer antigen for preventing,
XX PT treating, or ameliorating a medical condition, particular pancreatic
XX PT cancer, or for use in assays for diagnosing a pathological condition -
XX PS Claim 11; Page 1049; 1379pp; English.
XX CC C98773 to C99231 encode the human pancreatic cancer associated proteins,
XX CC called pancreatic cancer antigens, given in B54008 to B54466. The human
XX CC pancreatic cancer antigens have cytostatic, neuroprotective, nootropic,
XX CC immunomodulatory, relaxant, contraceptive, gynaecological, cardiant and
XX CC antiinflammatory activities, and can be used in gene therapy. The
XX CC polynucleotide and proteins can be used for preventing, treating, or
XX CC ameliorating a medical condition or in assays for diagnosing a
XX CC pathological condition or a susceptibility to one in a subject. Binding
XX CC partners to the proteins and the activity of the proteins can be
XX CC identified. The pancreatic cancer antigens can be used to detect, treat
XX CC or prevent pancreatic disorders, especially cancer. Agonists and
XX CC antagonists to the antigens can be screened for. The pancreatic cancer
XX CC antigen polynucleotides can be used to design nucleic acid hybridisation
XX CC probes that can be used in chromosome mapping, linkage analysis, tissue
XX CC identification and/or typing and a variety of forensic and diagnostic
XX CC methods. The proteins can be used to generate antibodies which are used
XX CC to purify, detect and target the polypeptides, including both in vivo
XX CC and in vitro diagnostic and therapeutic methods. The proteins can be
XX CC used to treat or prevent neural, immune system, muscular, reproductive,
XX CC gastrointestinal, pulmonary, cardiovascular, renal or proliferative
XX CC disorders. C99232 to C99240 and B54467 represent sequences used in the
XX CC exemplification of the present invention.
XX SO Sequence 77 AA;

Query Match 76.2%; Score 32; DB 21; Length 77;
Best Local Similarity 85.7%; Pred. No. 39;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GRLGVOD 7
   1111111
Db 59 grllvrd 65

```

Search completed: June 13, 2001, 14:14:36  
Job time: 377 sec

Wed Jun 13 14:59:36 2001

pct-us01-05825a-12.rag

Page 12

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

## OM protein - protein search, using sw model

Run on: June 13, 2001, 14:10:43 ; Search time 130.61 Seconds  
(without alignments)  
4.209 Million cell updates/sec

Title: PCT-US01-05825A-12

Perfect score: 42

Sequence: 1 GRGLVQDS 8

## Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 198801 seqs, 68722935 residues

Total number of hits satisfying chosen parameters: 198801

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 08  
Maximum Match 1008

Listing first 45 summaries

## Database :

PIR 67:\*  
1: PIR1:\*  
2: PIR2:\*  
3: PIR3:\*  
4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	37	88.1	540	2	T08662	probable signaling conserved hypothet
2	35	83.3	288	2	T48753	hypothetical prote
3	35	83.3	580	2	S76668	RNA polymerase II
4	34	81.0	176	2	B44457	DNA-directed RNA p
5	34	81.0	176	2	A44457	restriction precurs
6	34	81.0	1353	1	JH0675	nitrite hydratase
7	33	78.6	188	2	B37806	nitrite hydratase
8	33	78.6	207	1	S04472	nitrite hydratase
9	33	78.6	207	1	JN0705	hypothetical prote
10	33	78.6	295	2	H83989	hypothetical prote
11	33	78.6	378	2	T39466	alcohol dehydrogen
12	33	78.6	482	2	H69392	4-hydroxybutyrate
13	33	78.6	609	2	T00904	hypothetical prote
14	33	78.6	3436	2	S55659	legume protein 6
15	32	76.2	281	2	C72712	probable methionin
16	32	76.2	311	2	E69117	translation initia
17	32	76.2	350	1	JC4222	membrane dipeptida
18	32	76.2	512	2	T37819	probable zinc meta
19	32	76.2	686	2	T25987	hypothetical prote
20	32	76.2	925	2	T01384	hypothetical prote
21	32	76.2	1134	2	S31280	phytochrome - Mart
22	32	76.2	1158	2	T50454	probable rho1 GDP-
23	32	76.2	1236	2	E70977	hypothetical prote
24	31	73.8	199	2	T35977	probable acetyltra
25	31	73.8	278	2	G82370	shikimate 5-dehydr
26	31	73.8	526	2	B81533	glucose-6-phosphat
27	31	73.8	542	2	G64942	hypothetical prote
28	31	73.8	601	2	T32486	hypothetical prote
29	31	73.8	612	2	C83601	dihydroxy-acid deh

30	31	73.8	692	2	S47754	hypothetical prote
31	31	73.8	809	1	SYR0T	glutamine--crNA II
32	31	73.8	888	2	H65151	hypothetical 101.6
33	31	73.8	1596	2	T31338	cellulose synthase
34	30	71.4	216	2	H69443	conserved hypochet
35	30	71.4	262	2	S65988	yybi protein - Bac
36	30	71.4	278	2	T46587	hypothetical prote
37	30	71.4	277	2	T04441	hypothetical prote
38	30	71.4	285	2	T34836	probable dehydrata
39	30	71.4	361	2	I39529	hyprl protein - Al
40	30	71.4	363	2	H81410	hydrogenase isoenz
41	30	71.4	366	2	G82300	conserved hypochet
42	30	71.4	369	1	D64763	alcohol dehydrogen
43	30	71.4	369	1	S57525	alcohol dehydrogen
44	30	71.4	370	2	B83191	alcohol dehydrogen
45	30	71.4	373	1	A34119	alcohol dehydrogen

## ALIGNMENTS

RESULT 1  
T08662  
Probable signaling mediator DKF2p547G110.1 - human (fragment)  
C:Species: Homo sapiens (man)  
C>Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 24-Nov-1999  
C:Accession: T08662  
R:Poustka, A.; Klein, M.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.  
Submitted to the Protein Sequence Database, May 1999  
A:Reference number: Z16466  
A:Accession: T08662  
A:Molecule type: mRNA  
A:Residues: 1-540 <PDU>  
A:Cross-references: EMBL:AL049924  
A:Experimental source: fetal brain; clone DKF2p547G110  
C:Genetics:  
A:Note: DKF2p547G110.1  
C:Superfamily: SH2 homology  
C:Keywords: phosphoprotein  
F:326-334/Region: proline-rich  
F:384-478/Domain: SH2 homology <SH2>  
F:11,14,18/Binding site: phosphate (Ser) (covalent) #status predicted  
F:296,351,481/Binding site: phosphate (Tyr) (covalent) #status predicted

Query Match 88.1%; Score 37; DB 2; Length 540;  
Best Local Similarity 87.5%; Pred. No. 6.4;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GRGLVQDS 8  
Db 67 GRGLVSDG 74

RESULT 2  
T48753  
conserved hypothetical protein [Imported] - Neurospora crassa  
N:Alternate names: protein 13E11.10  
C:Species: Neurospora crassa  
C>Date: 05-May-2000 #sequence\_revision 05-May-2000 #text\_change 05-May-2000  
R:Schulte, U.; Aign, V.; Hohnsels, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatu  
Submitted to the Protein Sequence Database, April 2000  
A:Reference number: Z24541  
A:Accession: T48753  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-288 <SCH>  
A:Cross-references: EMBL:AL353820; GSPDB:GN00112; NCSP:13E11.10  
A:Experimental source: cosmid contig 13E11; strain 74  
C:Genetics:  
A:Gene: NCSP:13E11.10  
A:Map position: 2



Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GRGLVODG 8  
111:111

Db 302 GRVGVODG 309

## RESULT 7

B37806  
nitrile hydratase (EC 4.2.1.84) alpha chain - Brevibacterium sp. (strain R312) (fragment)  
C:Species: Brevibacterium sp.  
C:Date: 14-Jun-1991 #sequence\_revision 14-Jun-1991 #text\_change 16-Jul-1999  
C:Accession: B37806  
R:Mayaux, J.F.; Cerbasi, E.; Soubrrier, F.; Faucher, D.; Petre, D.  
J. Bacteriol. 172, 6764-6773, 1990  
A:Title: Purification, cloning, and primary structure of an enantiomer-selective amidase  
A:Reference number: A37806; MUID:91072222  
A:Accession: B37806  
A:Molecule type: DNA  
A:Residues: 1-188 <NAV>  
A:Cross-references: GB:M60264; GB:M32282; NID:9144090; PIDN:AAA62722.1; PID:9144092  
C:Function:  
A:Description: catalyzes the reaction of one molecule of water and an aliphatic nitrile  
C:Superfamily: nitrile hydratase alpha chain  
C:Keywords: carbon-oxygen lyase; hydro-lyase; iron; metalloprotein  
F:2-188/Product: nitrile hydratase alpha chain #status predicted <MAT>  
F:110,113,114,115/Binding site: nitrosyl iron (Cys, Cys, Ser, Cys) (shared with beta cha  
F:113/Modified site: cysteine sulfenic acid (Cys) #status predicted  
F:115/Modified site: cysteine sulfenic acid (Cys) #status predicted

Query Match 78.6%; Score 33; DB 2; Length 188;  
Best Local Similarity 75.0%; Pred. No. 15;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GRGLVODG 8  
111:111

Db 32 GKGLVODG 39

## RESULT 8

S04472  
nitrile hydratase (EC 4.2.1.84) alpha chain [validated] - Rhodococcus sp.  
C:Species: Rhodococcus sp.  
C:Date: 21-Nov-1998 #sequence\_revision 21-Nov-1998 #text\_change 15-Sep-2000  
C:Accession: S04472; A30463; S16606; S02070; S54105; S62095; S15071; S28887  
R:Ikehata, O.; Nishiyama, M.; Horinouchi, S.; Beppu, T.  
Eur. J. Biochem. 181, 563-570, 1999  
A:Title: Primary structure of nitrile hydratase deduced from the nucleotide sequence of  
A:Reference number: S04471; MUID:89276338  
A:Accession: S04472  
A:Molecule type: DNA  
A:Residues: 1-207 <IKE1>  
A:Cross-references: EMBL:X14668; NID:946429; PIDN:CAA32797.1; PID:946430  
A:Experimental source: strain N-774  
A:Accession: A30463  
A:Molecule type: protein  
A:Residues: 2-19;47-66;144-159 <IKE2>  
R:Nishiyama, M.  
submitted to the EMBL Data Library, March 1989  
A:Reference number: S16606  
A:Accession: S16606  
A:Molecule type: DNA  
A:Residues: 1-17, 'A', 19-207 <NIS>  
A:Cross-references: EMBL:X14668  
R:Endo, T.; Watanabe, I.  
FEBS Lett. 243, 61-64, 1989  
A:Title: Nitrile hydratase of Rhodococcus sp. N-774. Purification and amino acid sequenc  
A:Reference number: S02070; MUID:89153549  
A:Accession: S02070  
A:Molecule type: protein  
A:Residues: 2-19, 'T' <END>  
R:Bigey, F.; Chebrou, H.; Arnaud, A.; Galzy, P.

submitted to the EMBL Data Library, March 1995  
A:Description: Cloning, sequencing of the modified nitrile hydratase gene from mutant  
A:Reference number: S54104  
A:Accession: S54105  
A:Molecule type: DNA  
A:Residues: 1-207 <BIG>  
A:Cross-references: EMBL:248769; NID:9769823; PIDN:CAA8685.1; PID:9769825  
A:Experimental source: strain ACV2  
R:Nishiyama, M.  
submitted to the EMBL Data Library, July 1990  
A:Reference number: S18291  
A:Accession: S62095  
A:Molecule type: DNA  
A:Residues: 1-207 <NI2>

A:Cross-references: EMBL:X54074; NID:946410; PIDN:CAA38010.1; PID:946412  
A:Experimental source: strain N-774  
R:Hashimoto, Y.; Nishiyama, M.; Ikehata, O.; Horinouchi, S.; Beppu, T.  
Biochim. Biophys. Acta 1088, 225-233, 1991  
A:Title: Cloning and characterization of an amidase gene from Rhodococcus species N-7  
A:Reference number: S15070; MUID:91159474  
A:Accession: S15071  
A:Molecule type: DNA  
A:Residues: 1-29;30,197-207 <HNS>  
A:Cross-references: EMBL:X54074  
A:Experimental source: strain N-774  
R:Hung, W.; Schneider, G.; Lindqvist, Y.  
submitted to the Brookhaven Protein Data Bank, April 1997  
A:Reference number: A73039; PDB:1AHJ  
A:Contents: annotation: X-ray crystallography, 2.65 angstroms, residues 10-17, 'A', 19-  
R:Nagashima, S.; Nakasako, M.; Dojima, N.; Tsujimura, M.; Takio, K.; Odaka, M.; Yohad  
Nature Struct. Biol. 5, 347-351, 1998  
A:Title: Novel non-heme iron center of nitrile hydratase with a claw setting of oxyge  
A:Reference number: A58907; MUID:98246406  
A:Contents: annotation: X-ray crystallography, 1.7 angstroms; mass spectroscopic iden  
C:Comment: An activating protein (see PIR:JC2313) is required to generate the active  
C:Function:  
A:Description: catalyzes the reaction of one molecule of water and an aliphatic nitril  
C:Superfamily: nitrile hydratase alpha chain  
C:Keywords: carbon-oxygen lyase; hydro-lyase; iron; metalloprotein  
F:2-207/Product: nitrile hydratase alpha chain #status experimental  
F:110,113,114,115/Binding site: nitrosyl iron (Cys, Cys, Ser, Cys) (shared with beta  
F:113/Modified site: cysteine sulfenic acid (Cys) #status experimental  
F:115/Modified site: cysteine sulfenic acid (Cys) #status experimental

Query Match 78.6%; Score 33; DB 1; Length 207;  
Best Local Similarity 75.0%; Pred. No. 17;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GRGLVODG 8  
111:111

Db 32 GKGLVODG 39

## RESULT 9

JN0705  
nitrile hydratase (EC 4.2.1.84) alpha chain - Rhodococcus erythropolis  
C:Species: Rhodococcus erythropolis  
C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 16-Jun-2000  
C:Accession: JN0705  
R:Duran, R.; Nishiyama, M.; Horinouchi, S.; Beppu, T.  
Biosci. Biotechnol. Biochem. 57, 1323-1328, 1993  
A:Title: Characterization of nitrile hydratase genes cloned by DNA screening from Rhod  
A:Reference number: PN0570; MUID:93379358  
A:Accession: JN0705  
A:Molecule type: DNA  
A:Residues: 1-207 <DNR>  
A:Cross-references: DDBJ:D14454; NID:9441203; PIDN:BA03348.1; PID:9559379  
A:Experimental source: strain JCM6823  
C:Genetics:  
A:Gene: NHase  
C:Function:  
A:Description: catalyzes the reaction of one molecule of water and an aliphatic nitril

C:Superfamily: nitrile hydratase alpha chain.  
 C:Keywords: carbon-oxygen lyase; hydro-lyase; iron; metalloprotein  
 F:2-207/Product: nitrile hydratase alpha chain #status predicted <MAY>  
 F:110,113,114,115/Binding site: nitrosyl iron (Cys, Cys, Ser, Cys) (shared with beta cha  
 F:115/Modified site: cysteine sulfenic acid (Cys) #status predicted  
 F:115/Modified site: cysteine sulfenic acid (Cys) #status predicted

Query Match  
 Best Local Similarity 78.6%; Score 33; DB 1; Length 207;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 GRGLVQDG 8  
 |||||  
 Db 32 GRGLVQDG 39

RESULT 10  
 H83989  
 hypothetical protein BH2720 [imported] - Bacillus halodurans (strain C-125)  
 C:Species: Bacillus halodurans  
 C:Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 31-Dec-2000  
 C:Accession: H83989  
 R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fuji, F.; Hirai  
 Nucleic Acids Res. 28, 4317-4331, 2000  
 A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and  
 A:Reference number: A83650; MUID:20263314  
 A:Accession: H83989  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-295 <STO>  
 A:Cross-references: GB:AP001516; GB:BA000004; NID:910175192; PIDN:BA06439.1; GSPDB:GN00  
 A:Experimental source: strain C-125  
 C:Genetics:  
 A:Gene: BH2720

Query Match  
 Best Local Similarity 87.5%; Score 33; DB 2; Length 295;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GRGLVQDG 8  
 |||||  
 Db 220 GRGLVQDG 227

RESULT 11  
 T39466  
 alcohol dehydrogenase (EC 1.1.1.1) class III [similarity] - fission yeast (Schizosacchar  
 N:Contains: formaldehyde dehydrogenase (glutathione) (EC 1.2.1.1)  
 C:Species: Schizosaccharomyces pombe  
 C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 08-Dec-2000  
 C:Accession: T39466; T43041  
 R:Mc Dougal, R.; Rajandream, M.A.; Barrell, B.G.; Seeger, K.; Quail, M.; Harris, D.  
 submitted to the EMBL Data Library, July 1999  
 A:Reference number: 221856  
 A:Accession: T39466  
 A:Status: translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-378 <MCD>  
 A:Cross-references: EMBL:AL096874; PIDN:CAB51339.1; GSPDB:GN00067; SPDB:SPBC1539.07c  
 A:Experimental source: strain 972h-; cosmid c1539  
 R:Ishihaka, S.; Kato, K.; Nakai, K.; Okayama, H.; Nojima, H.  
 DNA Res. 4, 363-369, 1997  
 A:Title: Identification of open reading frames in Schizosaccharomyces pombe cDNAs.  
 A:Reference number: Z17323; MUID:98162722  
 A:Accession: T43041  
 A:Status: translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-126, 'S', 207-378 <YOS>  
 A:Cross-references: EMBL:D89220; NID:91749647; PIDN:BA13881.1; PID:dl014578; PID:917496  
 C:Genetics:  
 A:Experimental source: strain PR45

A:Gene: SPDB:SPBC1539.07c  
 A:Map position: 2  
 A:Introns: 11/3; 26/2  
 C:Function:

A:Description: catalyzes the oxidation of primary and secondary alcohols to aldehydes  
 C:Superfamily: alcohol dehydrogenase  
 C:Keywords: alcohol metabolism; NAD; oxidoreductase; zinc  
 F:47,69,176/Binding site: zinc, catalytic (Cys, His, Cys) #status predicted

Query Match  
 Best Local Similarity 78.6%; Score 33; DB 2; Length 378;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 GRGLVQDG 8  
 |||||  
 Db 121 GRGLVQDG 128

RESULT 12  
 H69392  
 4-hydroxybutyrate CoA transferase (cat2-1) homolog - Archaeoglobus fulgidus  
 C:Species: Archaeoglobus fulgidus  
 C:Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 17-Mar-2000  
 C:Accession: H69392  
 R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dod  
 , Pletschmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E  
 Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.  
 Nature 390, 364-370, 1997  
 A:Authors: Overbeek, T.; Cotton, M.D.; Spriggs, T.; Attiach, P.; Kaine, B.P.; Sykes,  
 Smith, H.O.; Woese, C.R.; Venter, J.C.  
 A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing arch  
 A:Reference number: A69250; MUID:98049343  
 A:Accession: H69392  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-482 <KLE>  
 A:Cross-references: GB:AF001025; GB:AE000782; NID:92689348; PIDN:AMB90101.1; PID:9264  
 C:Superfamily: acetyl-CoA hydrolase

Query Match  
 Best Local Similarity 78.6%; Score 33; DB 2; Length 482;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 GRGLVQDG 8  
 |||||  
 Db 225 GRGLVQDG 232

RESULT 13  
 T00904  
 hypothetical protein F21B7.18 - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 12-Feb-1999 #sequence\_revision 12-Feb-1999 #text\_change 22-Oct-1999  
 C:Accession: T00904  
 R:Shim, P.; Buehler, E.; Dewar, K.; Feng, J.; Kim, C.; Li, Y.; Sun, H.; Conway, A.;  
 submitted to the EMBL Data Library, January 1998  
 A:Description: Genomic sequence for Arabidopsis thaliana BAC F21B7.  
 A:Reference number: Z14208  
 A:Accession: T00904  
 A:Status: translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-609 <SHI>  
 A:Cross-references: EMBL:AC002560; NID:92618677; PID:92809249; GSPDB:GN00059; ATSP:F2  
 C:Genetics:  
 A:Gene: ATSP:F21B7.18  
 A:Map position: 1

Query Match  
 78.6%; Score 33; DB 2; Length 609;

Best Local Similarity 75.0%; Pred. No. 52;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 GRGLVODG 8  
|||  
Db 258 GRGLVODG 265

## RESULT 14

S55659  
legument protein 64 - equine herpesvirus 2  
C:Species: equine herpesvirus 2  
C>Date: 27-Oct-1995 #sequence\_revision 03-Nov-1995 #text\_change 08-Oct-1999  
C:Accession: S55659  
R:Jefford, E.A.R.; Watson, M.S.; Alrd, H.C.; Perry, J.; Davison, A.J.  
J. Mol. Biol. 249, 520-528, 1995  
A:Title: The DNA sequence of equine herpesvirus 2.  
A:Reference number: S55594; MUID:95302501  
A:Accession: S55659  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-3436 <TEL>  
A:Cross-references: GB:U20824; NID:9695172; PIDN:AA13852.1; PID:9695237  
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, February 1995

Query Match 78.6%; Score 33; DB 2; Length 3436;  
Best Local Similarity 75.0%; Pred. No. 3.2e+02;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 GRGLVODG 8  
|||  
Db 1048 GRGLVODG 1055

## RESULT 15

C72712  
probable methionine aminopeptidase APE114 - Aeropyrum pernix (strain K1)  
C:Species: Aeropyrum pernix  
C>Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 20-Jun-2000  
R:Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Hatakeyama, Y.; Jinno, K.; Takahara, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudo, Y.; Yamazaki, J.; K  
DNA Res. 6, 83-101, 1999  
A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyr  
A:Reference number: A72450; MUID:9310339  
A:Accession: C72712  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-281 <KAM>  
A:Cross-references: DDBJ:AP000060; NID:95104188; PIDN:BA80099.1; PID:95104784  
A:Experimental source: strain K1  
C:Genetics:  
A:Gene: APE114  
C:Superfamily: Escherichia coli methionyl aminopeptidase

Query Match 76.2%; Score 32; DB 2; Length 281;  
Best Local Similarity 62.5%; Pred. No. 37;  
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 1 GRGLVODG 8  
|||  
Db 171 GRGLVODG 178

Search completed: June 13, 2001, 14:10:44  
Job time: 145 sec

*This Page Blank (uspto)*



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OW protein - protein search, using sw model

Run on: June 13, 2001, 14:21:44 ; Search time 74.44 Seconds  
(without alignments)  
3.681 Million cell updates/sec

Title: PCT-US01-05825A-12

Perfect score: 42

Sequence: 1 GRGLVODG 8

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_39:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	34	81.0	176	1	RPB7_ARATH
2	34	81.0	176	1	RPB7_SOYAN
3	33	78.6	206	1	NHAA_RHOER
4	32	76.2	311	1	E2B1_METTH
5	32	76.2	512	1	YE94_SCHRO
6	32	76.2	1134	1	PHY1_SELMA
7	31	73.8	809	1	SYO_YEAST
8	31	73.8	888	1	YH04_ECOLI
9	30.5	72.6	344	1	PROA_SPIRX
10	30	71.4	216	1	KPR2_ARCFU
11	30	71.4	262	1	YBBI_BACSU
12	30	71.4	361	1	HYPB_ALCEU
13	30	71.4	369	1	ADH3_ECOLI
14	30	71.4	373	1	ADH3_HORSE
15	30	71.4	373	1	ADH3_HUMAN
16	30	71.4	373	1	ADH3_MOUSE
17	30	71.4	373	1	ADH3_RABIT
18	30	71.4	373	1	ADH3_RAT
19	30	71.4	374	1	ADH7_HUMAN
20	30	71.4	374	1	ADH7_RAT
21	30	71.4	379	1	CYR6_MOUSE
22	30	71.4	381	1	CYR6_HUMAN
23	30	71.4	434	1	PEL_TLULO
24	30	71.4	518	1	LEU1_BACSU
25	30	71.4	534	1	EPAL_MOUSE
26	30	71.4	609	1	LAC1_EMENT
27	30	71.4	618	1	KEXF_HAERI
28	30	71.4	769	1	VPL_BPPH6
29	30	71.4	770	1	SM4F_HUMAN
30	30	71.4	810	1	NEL1_HUMAN
31	30	71.4	810	1	NEL1_RAT
32	30	71.4	1018	1	YK26_CAEEL
33	30	71.4	1527	1	CAIH_MOUSE

34	30	71.4	2264	1	POL1_TBRSV
35	29	69.0	200	1	YE15_HAERI
36	29	69.0	375	1	ADH1_PARDE
37	29	69.0	376	1	ADH1_RHOSH
38	29	69.0	389	1	ATVY_RHIME
39	29	69.0	439	1	G6PD_CHLTR
40	29	69.0	483	1	MURE_CHLTP
41	29	69.0	578	1	VAT4_METHA
42	29	69.0	617	1	PSY1_PSEAE
43	29	69.0	651	1	NADO_THERB
44	29	69.0	662	1	YAE8_YEAST
45	29	69.0	673	1	ESR2_MICUN

## ALIGNMENTS

RESULT 1	
ID RPB7_ARATH	STANDARD; PRT; 176 AA.
AC P38421:	
DT 01-OCT-1994 (Rel. 30, Created)	
DT 01-OCT-1994 (Rel. 30, Last sequence update)	
DT 01-OCT-2000 (Rel. 40, Last annotation update)	
DE DNA-DIRECTED RNA POLYMERASE II 19 KDA POLYPEPTIDE (EC 2.7.7.6)	
DE (RNA POLYMERASE II SUBUNIT 5).	
GN RPB19.	
OS Arabidopsis thaliana (Mouse-ear cress).	
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;	
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;	
OC Brassicales; Brassicaceae; Arabidopsids.	
OX NCBI_Taxid=3702;	
RN [1]	
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.	
RC STRAIN=CV. COLUMBIA;	
RX MEDLINE=93054645; PubMed=1429663;	
RA Ulmasov T., Gullfoyle T.J.;	
RT "Sequence of the fifth largest subunit of RNA polymerase II from	
RT plants."	
RL J. Biol. Chem. 267:23165-23169(1992).	
CC - FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION	
CC OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS	
CC SUBSTRATES.	
CC - CATALYTIC ACTIVITY: N NUCLEOSIDE TRIPHOSPHATE - N PYROPHOSPHATE +	
CC RNA(N).	
CC - SUBUNIT: EACH CLASS OF RNA POLYMERASE IS ASSEMBLED FROM 9 TO 14	
CC DIFFERENT POLYPEPTIDES. THIS SUBUNIT IS THE FIFTH LARGEST	
CC COMPONENT OF RNA POLYMERASE II.	
CC - SUBCELLULAR LOCATION: NUCLEAR.	
CC - MISCELLANEOUS: THREE DISTINCT ZINC-CONTAINING RNA POLYMERASES ARE	
CC FOUND IN EUKARYOTIC NUCLEI: POLYMERASE I FOR THE RIBOSOMAL RNA	
CC PRECURSOR, POLYMERASE II FOR THE MRNA PRECURSOR, AND POLYMERASE	
CC III FOR 5S AND TRNA GENES.	
CC -----	
CC This SWISS-PROT entry is copyright. It is produced through a collaboration	
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -	
CC the European Bioinformatics Institute. There are no restrictions on its	
CC use by non-profit institutions as long as its content is in no way	
CC modified and this statement is not removed. Usage by and for commercial	
CC entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>	
CC or send an email to <a href="mailto:license@sib-sib.ch">license@sib-sib.ch</a> ).	
CC -----	
DR EMBL: M90505; AAA32861.1; -	
DR PIR: A44457; A44457.	
DR Mende1: 7052; ARATH:Rpb19.1.	
DR InterPro: IPR003029; -	
DR Pfam: PF00575; S1.1.	
KW Transferase: DNA-directed RNA polymerase; Transcription;	
KW Nuclear protein.	
SQ SEQUENCE 176 AA; 19462 MW; A508334EA3EAB186 CRC64;	

Query Match

81.0%; Score 34; DB 1; Length 176;

Best Local Similarity 62.5%; Pred. No. 6.3;  
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 GRGIVDGG 8  
1:|||||  
Db 55 GKGLIRDG 62

## RESULT 2

RPP7\_SOYBN

ID RPP7\_SOYBN STANDARD: PRT; 176 AA.  
AC P46279;

DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)

DE (RNA-DIRECTED RNA POLYMERASE II SUBUNIT 5).  
DE (RNA POLYMERASE II SUBUNIT 5).

OS Glycine max (Soybean).  
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;

OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eustosids I;  
OC Fabales; Fabaceae; Papilionoideae; Glycine.

OX NCBI\_Taxid=38477;  
RN [1]

RP SEQUENCE FROM N.A.  
RX MEDLINE-93054645; Pubmed-1429663;  
RA Umasov T.N., Guilfoyle T.J.;

RT "Sequence of the fifth largest subunit of RNA polymerase II from  
plants.";

RL J. Biol. Chem. 267:23165-23169(1992).  
CC -1- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION  
OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS  
SUBSTRATES.

CC -1- CATALYTIC ACTIVITY: N NUCLEOSIDE TRIPHOSPHATE = N PYROPHOSPHATE +  
RNA(N).

CC -1- SUBUNIT: EACH CLASS OF RNA POLYMERASE IS ASSEMBLED FROM 9 TO 14  
DIFFERENT POLYPEPTIDES. THIS SUBUNIT IS THE FIFTH LARGEST  
COMPONENT OF RNA POLYMERASE II.

CC -1- MISCELLANEOUS: LOCATION: NUCLEAR.  
CC -1- SUBCELLULAR LOCATION: NUCLEAR.

CC -1- FOUND IN EUKARYOTIC NUCLEI: POLYMERASE I FOR THE RIBOSOMAL RNA  
PRECURSOR, POLYMERASE II FOR THE MRNA PRECURSOR, AND POLYMERASE  
III FOR 5S AND TRNA GENES.

CC -1- THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -1- EMBL: M90504; AAA34005.1; -  
DR InterPro: IPR003029; -  
DR Pfam: PF00575; S1.1.  
KW Transferrase; DNA-directed RNA polymerase; Transcription;  
KW Nuclear protein.

CC SEQUENCE 176 AA; 19625 MW; FCCP9AAB096BB36 CRC64;

Query Match 81.0%; Score 34; DB 1; Length 176;  
Best Local Similarity 62.5%; Pred No. 6.3;  
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 GRGIVDGG 8  
1:|||||  
Db 55 GKGLIRDG 62

RESULT 3  
ID NHAA\_RHOER STANDARD: PRT; 206 AA.  
AC P13448;  
DT 01-JAN-1990 (Rel. 13, Created)

DT 01-AUG-1991 (Rel. 19, Last sequence update)  
DE 15-JUL-1999 (Rel. 38, Last annotation update)  
DE NITRILE HYDRATASE SUBUNIT ALPHA (EC 4.2.1.84) (NITRILASE) (NHASE).  
GN NTHA.

OS Rhodococcus erythropolis.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Corynebacteriaceae; Nocardiaceae; Rhodococcus.

OX NCBI\_Taxid=1833;  
RN [1]

RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
RC STRAIN-N-774;  
RX MEDLINE-89276338; Pubmed-2659343;  
RA Ikehata O., Nishiyama M., Horinouchi S., Beppu T.;

RT "Primary structure of nitrile hydratase deduced from the nucleotide  
sequence of a Rhodococcus species and its expression in Escherichia  
coli.";

RL Eur. J. Biochem. 181:563-570(1989).  
CC -1- FUNCTION: NHASE CATALYZES THE HYDRATION OF VARIOUS NITRILE  
AMIDES TO THE CORRESPONDING AMIDES. INDUSTRIAL PRODUCTION OF  
ACRYLAMIDE IS NOW BEING DEVELOPED USING SOME OF THESE ENZYMES.

CC -1- CATALYTIC ACTIVITY: AN ALIPHATIC AMIDE = A NITRILE + H(2O).  
CC -1- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA-CHAIN.  
CC -1- SIMILARITY: TO OTHER NITRILE HYDRATASES SUBUNIT ALPHA.

CC -1- THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -

DT 01-AUG-1991 (Rel. 19, Last sequence update)  
DE 15-JUL-1999 (Rel. 38, Last annotation update)  
DE NITRILE HYDRATASE SUBUNIT ALPHA (EC 4.2.1.84) (NITRILASE) (NHASE).  
GN NTHA.

OS Rhodococcus erythropolis.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Corynebacteriaceae; Nocardiaceae; Rhodococcus.

OX NCBI\_Taxid=1833;  
RN [1]

RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
RC STRAIN-N-774;  
RX MEDLINE-89276338; Pubmed-2659343;  
RA Ikehata O., Nishiyama M., Horinouchi S., Beppu T.;

RT "Primary structure of nitrile hydratase deduced from the nucleotide  
sequence of a Rhodococcus species and its expression in Escherichia  
coli.";

RL Eur. J. Biochem. 181:563-570(1989).  
CC -1- FUNCTION: NHASE CATALYZES THE HYDRATION OF VARIOUS NITRILE  
AMIDES TO THE CORRESPONDING AMIDES. INDUSTRIAL PRODUCTION OF  
ACRYLAMIDE IS NOW BEING DEVELOPED USING SOME OF THESE ENZYMES.

CC -1- CATALYTIC ACTIVITY: AN ALIPHATIC AMIDE = A NITRILE + H(2O).  
CC -1- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA-CHAIN.  
CC -1- SIMILARITY: TO OTHER NITRILE HYDRATASES SUBUNIT ALPHA.

CC -1- THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -

DT 01-AUG-1991 (Rel. 19, Last sequence update)  
DE 15-JUL-1999 (Rel. 38, Last annotation update)  
DE NITRILE HYDRATASE SUBUNIT ALPHA (EC 4.2.1.84) (NITRILASE) (NHASE).  
GN NTHA.

OS Rhodococcus erythropolis.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Corynebacteriaceae; Nocardiaceae; Rhodococcus.

OX NCBI\_Taxid=1833;  
RN [1]

RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
RC STRAIN-N-774;  
RX MEDLINE-89276338; Pubmed-2659343;  
RA Ikehata O., Nishiyama M., Horinouchi S., Beppu T.;

RT "Primary structure of nitrile hydratase deduced from the nucleotide  
sequence of a Rhodococcus species and its expression in Escherichia  
coli.";

RL Eur. J. Biochem. 181:563-570(1989).  
CC -1- FUNCTION: NHASE CATALYZES THE HYDRATION OF VARIOUS NITRILE  
AMIDES TO THE CORRESPONDING AMIDES. INDUSTRIAL PRODUCTION OF  
ACRYLAMIDE IS NOW BEING DEVELOPED USING SOME OF THESE ENZYMES.

CC -1- CATALYTIC ACTIVITY: AN ALIPHATIC AMIDE = A NITRILE + H(2O).  
CC -1- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA-CHAIN.  
CC -1- SIMILARITY: TO OTHER NITRILE HYDRATASES SUBUNIT ALPHA.

CC -1- THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -

DT 01-AUG-1991 (Rel. 19, Last sequence update)  
DE 15-JUL-1999 (Rel. 38, Last annotation update)  
DE NITRILE HYDRATASE SUBUNIT ALPHA (EC 4.2.1.84) (NITRILASE) (NHASE).  
GN NTHA.

OS Rhodococcus erythropolis.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Corynebacteriaceae; Nocardiaceae; Rhodococcus.

OX NCBI\_Taxid=1833;  
RN [1]

RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
RC STRAIN-N-774;  
RX MEDLINE-89276338; Pubmed-2659343;  
RA Ikehata O., Nishiyama M., Horinouchi S., Beppu T.;

RT "Primary structure of nitrile hydratase deduced from the nucleotide  
sequence of a Rhodococcus species and its expression in Escherichia  
coli.";

RL Eur. J. Biochem. 181:563-570(1989).  
CC -1- FUNCTION: NHASE CATALYZES THE HYDRATION OF VARIOUS NITRILE  
AMIDES TO THE CORRESPONDING AMIDES. INDUSTRIAL PRODUCTION OF  
ACRYLAMIDE IS NOW BEING DEVELOPED USING SOME OF THESE ENZYMES.

CC -1- CATALYTIC ACTIVITY: AN ALIPHATIC AMIDE = A NITRILE + H(2O).  
CC -1- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA-CHAIN.  
CC -1- SIMILARITY: TO OTHER NITRILE HYDRATASES SUBUNIT ALPHA.

CC -1- THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: X14668; CAA33797.1; -  
 DR EMBL: X54074; CAA38010.1; -  
 DR EMBL: Z48769; CAA88685.1; -  
 DR EMBL: M60264; AAA62722.1; -  
 DR PIR: S04472; S04472.  
 DR PIR: B37806; B37806.  
 DR PDB: 1AHJ; 08-APR-98.  
 DR PDB: 2AHJ; 16-FEB-99.  
 KW Lyase; Iron; 3D-structure.  
 FT INIT\_MET 0  
 FT METAL 109 109 IRON.  
 FT METAL 112 112 IRON.  
 FT METAL 113 113 IRON.  
 FT METAL 114 114 IRON.  
 FT CONFLICT 17 17 P->A (IN REF. 1).  
 FT CONFLICT 19 19 S->T (IN REF. 4).  
 SQ SEQUENCE 206 AA; 22865 MM; 19C16FA01653E002 CRC64;

Query Match 78.6%; Score 33; DB 1; Length 206;  
 Best Local Similarity 75.0%; Pred. NO. 12;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Oy 1 GRLVODG 8  
 Db 31 GKGLVPDG 38

RESULT 4  
 ID E2B1\_METH STANDARD; PRT; 311 AA.  
 AC 027900;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE PUTATIVE TRANSLATION INITIATION FACTOR EIF-2B SUBUNIT 1 (EIF-2B  
 DE GDP-GTP EXCHANGE FACTOR).  
 GN MTH1872.  
 OS Methanobacterium thermoautotrophicum.  
 OC Archaea; Euryarchaeota; Methanobacteriales; Methanobacteriaceae;  
 CC Methanothermobacter.  
 OX NCBI\_TaxID=145262;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-DELTA H;  
 RX MEDLINE=98037514; PubMed=9371463;  
 RA Smith D.R., Doucette-Stamm L.A., Delonghery C., Lee H.-M., Dubois J.,  
 RA Aldredge T., Baahirzadeh R., Blakely D., Cook R., Gilbert K.,  
 RA Harrison D., Hoang L., Keagle P., Lamm W., Pochler B., Qiu D.,  
 RA Spadatore R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,  
 RA Uwan N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,  
 RA McDougall S., Shimer G., Goyal A., Pietrowski S., Church G.M.,  
 RA Daniels C.J., Mao J.-I., Rice P., Nolling J., Reeve J.N.;  
 RT "Complete genome sequence of Methanobacterium thermoautotrophicum  
 RT deltain: functional analysis and comparative genomics.";  
 RL J. Bacteriol. 179:7135-7155(1997).  
 CC -1- FUNCTION: CATALYZES THE EXCHANGE OF INITIATION FACTOR 2-BOUND GDP  
 CC FOR GTP (BY SIMILARITY).  
 CC -1- SUBUNIT: COMPLEX OF TWO DIFFERENT SUBUNITS (POTENTIAL).  
 CC -1- SIMILARITY: BELONGS TO THE EIF-2B ALPHA/BETA/DELTA SUBUNITS  
 CC FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: AE000939; AAB86338.1; -  
 DR InterPro: IPR000649; -  
 DR Pfam: PF01008; IF-2B; 1.  
 KW Hypothetical protein; Initiation factor; Protein biosynthesis.  
 SQ SEQUENCE 311 AA; 33871 MM; 797777EE6E321AF CRC64;

Query Match 76.2%; Score 32; DB 1; Length 311;  
 Best Local Similarity 62.5%; Pred. NO. 28;  
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Oy 1 GRLVODG 8  
 Db 121 GAGLIEDG 128

RESULT 5  
 ID YE94\_SCHPO STANDARD; PRT; 512 AA.  
 AC 013766;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE PROBABLE ZINC METALLOPEPTIDASE C17A5.04C PRECURSOR (RC 3.4.24.-).  
 GN SPAC17A5.04C.  
 OS Schizosaccharomyces pombe (fission yeast).  
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
 CC Schizosaccharomycetales; Schizosaccharomycetaceae;  
 CC Schizosaccharomycetes.  
 OX NCBI\_TaxID=4896;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=972;  
 RA Devlin K., Churcher C.M., Barrell B.G., Rajandream M.A., Wood V.;  
 RL submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: MAY ACT AS AN ATP-DEPENDENT ZINC METALLOPEPTIDASE  
 CC (BY SIMILARITY).  
 CC -1- COFACTOR: BINDS ONE ZINC ION (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B (ZINC  
 CC METALLOPROTEASE); ALSO KNOWN AS THE REPOLYSIN SUBFAMILY.  
 CC -1- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: Z98849; CAB1504.1; -  
 DR HSPD: P18619; IEVL.  
 DR MEROPS: M12.180; -  
 DR InterPro: IPR000130; -  
 DR InterPro: IPR001590; -  
 DR InterPro: IPR001762; -  
 DR Pfam: PF01421; Repolysin; 1.  
 DR Pfam: PF00200; disintegrin; 1.  
 DR PROSITE: PS50215; ADAM\_MEROPS; 1.  
 DR PROSITE: PS00427; DISINTEGRIN\_1; FALSE, NEG.  
 DR PROSITE: PS50214; DISINTEGRIN\_2; 1.  
 DR PROSITE: PS00142; ZINC\_PROTEASE; FALSE, NEG.  
 KW Hydrolyase; Metalloprotease; zinc; Metal-binding; Signal.  
 FT SIGNAL 1 15  
 FT CHAIN 16 512  
 FT METAL 229 229  
 FT ACET\_SITE 230 230  
 FT METAL 233 233  
 FT METAL 239 239

POTENTIAL.  
 PROBABLE ZINC METALLOPEPTIDASE C17A5.04C.  
 ZINC (CATALYTIC) (BY SIMILARITY).  
 BY SIMILARITY.  
 ZINC (CATALYTIC) (BY SIMILARITY).  
 ZINC (CATALYTIC) (PROBABLE).

SO SEQUENCE 512 AA; 56439 MW; 804BDA233621DAB CRC64;

Query Match  
Best Local Similarity 76.2%; Score 32; DB 1; Length 512;  
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 GRLVODG 8  
Db 319 GRLVODG 326

RESULT 6  
PHYL\_SELMA

ID PHYL\_SELMA STANDARD; PRT; 1134 AA.

AC Q01549;  
DT 01-JUL-1993 (Rel. 26, Created)  
DT 01-JUL-1993 (Rel. 26, Last sequence update)  
DE 15-JUL-1999 (Rel. 38, Last annotation update)  
GN PHYTOCHROME 1.  
OS PHY1.

OC Selaginella martensii (Martens's spike moss).  
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Lycopodiophyta;  
OC Isoetes; Selaginellales; Selaginellaceae; Selaginella.

OX NCBI\_TaxID=3247;  
RN [1]  
RP SEQUENCE FROM N.A.

RC STRAIN-SPRING;  
RA MEDLINE=93117303; PubMed=1475321;  
RA Hanelt S., Braun B., Marx S., Schneider-Poetsch H.A.W.;

RT "Phytochrome evolution: a phylogenetic tree with the first complete  
sequence of phytochrome from a cryptogamic plant (Selaginella  
Phlochem. Photobiol. 56:751-758(1992)).  
RT Martensii spring.";  
RL Photochem. Photobiol. 56:751-758(1992).

CC -1- FUNCTION: REGULATORY PHOTORECEPTOR WHICH EXISTS IN TWO FORMS THAT  
ARE REVERSIBLY INTERCONVERTIBLE BY LIGHT: THE PR FORM THAT ABSORBS  
MAXIMALLY IN THE RED REGION OF THE SPECTRUM AND THE PFR FORM THAT  
CC ABSORBS MAXIMALLY IN THE FAR-RED REGION. PHOTOCVERSION OF PR IN  
PFR INDUCES AN ARRAY OF MORPHOGENIC RESPONSES, WHEREAS  
CC RECONVERSION OF PFR TO PR CANCELS THE INDUCTION OF THOSE  
RESPONSES. PFR CONTROLS THE EXPRESSION OF A NUMBER OF NUCLEAR  
CC GENES INCLUDING THOSE ENCODING THE SMALL SUBUNIT OF RIBULOSE-  
BISPHOSPHATE CARBOXYLASE, CHLOROPHYLL A/B BINDING PROTEIN,  
CC PROTHIOPHYLLIDE REDUCTASE, RNA, ETC. IT ALSO CONTROLS  
THE EXPRESSION OF ITS OWN GENE(S) IN A NEGATIVE FEEDBACK FASHION.

CC -1- SUBUNIT: HOMODIMER.  
CC -1- PPM: CONTAINS ONE COVALENTLY LINKED TETRAPYRROLE CHROMOPHORE.  
CC -1- SIMILARITY: BELONGS TO THE PHYTOCHROME FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL Outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: X61458; CAA43698.1; -  
DR PIR: S31280; S31280.

DR InterPro: IPR000014; -  
DR InterPro: IPR000410; -  
DR InterPro: IPR001294; -  
DR InterPro: IPR003018; -

DR Pfam: PF01590; GAF. 1.  
DR Pfam: PF00989; PAS. 2.  
DR Pfam: PF00360; phytochrome. 1.

DR Pfam: PF00512; signal. 1.  
DR PRINTS: PR01033; phytochrome. 1.

DR PROSITE: PS00243; PHYTOCHROME. 1.  
DR PROSITE: PS00046; PHYTOCHROME-2. 1.

DR Transcription regulation; Photoreceptor; Phytochrome; Chromophore;  
KW Multigene family.

FT BINDING 324 324  
SEQUENCE 324 324

CHROMOPHORE (BY SIMILARITY).

FT DOMAIN 352 358 POLY-GLY.  
SO SEQUENCE 1134 AA; 124706 MW; 42819B9FAACC398C CRC64;

Query Match  
Best Local Similarity 76.2%; Score 32; DB 1; Length 1134;  
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 GRLVODG 8  
Db 1076 GRLVODG 1083

RESULT 7  
SYO\_YEAST

ID SYO\_YEAST STANDARD; PRT; 809 AA.  
AC P13188; Q12005;  
DT 01-JAN-1990 (Rel. 13, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DE 01-OCT-2000 (Rel. 40, Last annotation update)  
GN GLUTAMINYL-TRNA SYNTHETASE (EC 6.1.1.18) (GLUTAMINE--TRNA LIGASE)

DE (GLNRS).  
GN GLN4 OR YOR169W OR O3601.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.

OX NCBI\_TaxID=4932;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA MEDLINE=87280149; PubMed=3301841;  
RA Ludmerer S.W., Schlammel P.;

RT "Gene for yeast glutamine tRNA synthetase encodes a large  
signature sequence for a group of the aminoacyl-tRNA synthetases.";  
RT J. Biol. Chem. 262:10801-10806(1987).

RL [2]  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-S288C / FY1678;  
RX MEDLINE=97127828; PubMed=8972579;  
RA Madanla A., Poch O., Tarassov I.A., Winsor B., Martin R.P.;

RT "Analysis of a 22,956 bp region on the right arm of Saccharomyces  
cerevisiae chromosome XV.";  
RT Yeast 12:1153-1173(1996).

CC -1- CATALYTIC ACTIVITY: ATP + L-GLUTAMINE + TRNA(GLN) = AMP +  
PYROPHOSPHATE + L-GLUTAMINYL-TRNA(GLN).  
CC -1- SIMILARITY: BELONGS TO CLASS-I AMINOACYL-TRNA SYNTHETASE FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL Outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: M29184; AAA3646.1; -  
DR EMBL: M29185; - NOT\_ANNOTATED\_CDS.  
DR EMBL: U55021; AAB47415.1; -  
DR EMBL: 275076; CAA9374.1; -

DR PIR: A28494; SYBYOT.  
DR HSSP: P00962; 10RU.  
DR SGD: S0005694; GLN4.

DR InterPro: IPR000924; -  
DR InterPro: IPR001412; -  
DR Pfam: PF00746; tRNA-syn-1c. 1.

DR PROSITE: PS00178; AA-TRNA\_LIGASE\_I. 1.  
KW aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding.

FT SITE 258 268  
SITE 258 268  
FT SITE 495 499  
SITE 495 499

FT BINDING 498 498  
BINDING 498 498  
FT CONFLICT 179 179  
CONFLICT 179 179

SO SEQUENCE 809 AA; 93132 MW; C7AB13D02BCA83F6 CRC64;

Query Match 73.8%: Score 31; DB 1; Length 809;  
 Best Local Similarity 62.5%: Pred. No. 1.2e+02;  
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 GRCGLVODG 8  
 |||: |||  
 DB 356 GRCGLVODG 363

```

RESULT 8
YHJO_ECOLI STANDARD: PRT; 888 AA.
AC P37653; P37654; P76712; P76713;
DT 01-OCT-1994 (Rel. 30, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE HYPOTHETICAL 78.6 KDA PROTEIN IN DCTA-DPPF INTERGENIC REGION.
GN YHJO.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655.
RC MEDLINE=94316500; PubMed=8041620;
RA Soledad H., Burland V., Daniels D.L., Plunkett G. III, Blattner F.R.;
RT "Analysis of the Escherichia coli genome. V. DNA sequence of the
RT region from 76.0 to 81.5 minutes."
RL Nucleic Acids Res. 22:2576-2586(1994).
RN [2]
RP REVISIONS.
RC STRAIN-K12 / MG1655;
RC MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474(1997).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE
CC (POTENTIAL).
CC -1- SIMILARITY: TO ACETOBIACETYLINUM CELLULOSE SYNTHASE SUBUNIT
CC ACSA/BCSA.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U00039; AB181510.1; ALT_FRAME.
DR EMBL: U00039; CAB34650.1; ALT_FRAME.
DR EMBL: AB000430; AAC76558.1; -.
DR EcoGene: EG12260; ynfO.
DR InterPro: IPR001173; -.
DR Pfam: PF00535; Glycosyltransferase; Cellulose biosynthesis; Transferase;
KM Hypothetical protein; Transmembrane; Inner membrane.
KW Glycosyltransferase; Transmembrane; Inner membrane.
FT TRANSMEM 50 POTENTIAL.
FT TRANSMEM 345 POTENTIAL.
FT TRANSMEM 367 POTENTIAL.
FT TRANSMEM 412 POTENTIAL.
FT TRANSMEM 460 POTENTIAL.
FT TRANSMEM 488 POTENTIAL.
FT TRANSMEM 654 POTENTIAL.
FT TRANSMEM 674 POTENTIAL.
SQ SEQUENCE 888 AA; 101597 MW; C32E169E7E1DCC4F CRC64;

```

Best Local Similarity 100.0%: Pred. No. 1.3e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 GRCGLVODG 8  
 |||: |||  
 DB 433 GRCGLVODG 438

```

RESULT 9
RPOA_SPIRX STANDARD: PRT; 344 AA.
AC O98462;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE DNA-DIRECTED RNA POLYMERASE ALPHA CHAIN (EC 2.7.7.6) (TRANSCRIPTASE
DE ALPHA CHAIN) (RNA POLYMERASE ALPHA SUBUNIT).
GN RPOA.
OS Spirogyra maxima.
OC Chloroplast.
OC Eukaryota; Viridiplantae; Zygnemophyta; Zygnemophyceae; Zygnematales;
OC Zygnematales; Spirogyra.
OX NCBI_TaxID=3180;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-UTEX LB 2495;
RC Lee J., Manhart J.R.;
RT "Chloroplast rpl23 gene cluster of Spirogyra maxima (Charophyceae),
RT shared by land plants."
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION
CC OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS
CC SUBSTRATES.
CC -1- CATALYTIC ACTIVITY: N NUCLEOSIDE TRIPHOSPHATE - N PYROPHOSPHATE +
CC RNA(N).
CC -1- SUBUNIT: IN CHLOROPLAST THE RNA POLYMERASE IS COMPOSED OF FOUR
CC SUBUNITS: ALPHA, BETA, BETA', AND BETA".
CC -1- SIMILARITY: BELONGS TO THE RNA POLYMERASE ALPHA CHAIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF050665; AAC95318.1; -.
DR HSSP: P00574; 1COO.
DR InterPro: IPR001700; -.
DR Pfam: PF01000; RNA_Pol_A_bac; 1.
KW Transferase; DNA-directed RNA polymerase; Transcription; Chloroplast.
SQ SEQUENCE 344 AA; 38837 MW; 2B52BFF90446CB0E CRC64;

```

Query Match 72.6%: Score 30.5; DB 1; Length 344;  
 Best Local Similarity 77.8%: Pred. No. 63;  
 Matches 7; Conservative 1; Mismatches 0; Indels 1; Gaps 1;

OY 1 GRCGLVODG 8  
 |||: |||  
 DB 156 GRCGLVODG 164

```

RESULT 10
KPT2_ARCFU STANDARD: PRT; 216 AA.
ID KPT2_ARCFU
AC O28719;
DT 01-OCT-2000 (Rel. 40, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE PROBABLE RNA 2'-PHOSPHOTRANSFERASE 2 (EC 2.7.7.-).
GN KPTA2 OR AF1553.

```

Query Match 73.8%: Score 31; DB 1; Length 888;

```

OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobales; Archaeoglobaceae;
OC Archaeoglobus.
OX NCBI_TaxID=2234;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-VC-16 / DSM 4304 / ATCC 49558;
RX MEDLINE=98049343; PubMed=9389475;
RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
RA Fleischmann D.D., Kiehl J.A., Smith T.A., Peterson J.D.,
RA Kirschner S., Reich C.I., McNeil L.K., McKenney K., Adams M.D., Loftis S.,
RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,
RA Cotton M.D., Spriggs T., Artach P., Kaine B.P., Sykes S.M.,
RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA Venter J.C.;
RT "The complete genome sequence of the hyperthermophilic, sulphate-
RT reducing archaeon Archaeoglobus fulgidus."
RL Nature 390:364-370(1997).
CC -1- FUNCTION: REMOVES THE 2'-PHOSPHATE FROM RNA VIA AN INTERMEDIATE IN
CC WHICH THE PHOSPHATE IS ADP-RIBOSYLATED BY NAD FOLLOWED BY A
CC PRESUMED TRANSFER OF THE ADP-RIBOSYL GROUP TO THE RNA AND GENERATE ADP-
CC RIBOSE 1'-2'-CYCLIC PHOSPHATE (APPR>P). MAY FUNCTION AS AN ADP-
CC -1- SIMILARITY: BELONGS TO THE KPN / TPR1 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF000695; AAB89693.1; -
DR TIGR: AF1553; -
DR InterPro: IPR002745; -
DR Pfam: PF01885; D0F60; 1.
DR Transferrase; NAD.
SQ SEQUENCE 216 AA; 24825 MW; 893EFBA39EB2BC CRC64;

Query Match 71.4%; Score 30; DB 1; Length 216;
Best Local Similarity 71.4%; Pred. No. 50;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 2 GRLVODG 8
DB 174 KGLMDG 180

RESULT 11
YBBI_BACSU STANDARD: PRT; 262 AA.
AC P37495;
DT 01-OCT-1994 (Rel. 30. Created)
DT 01-OCT-1994 (Rel. 30. Last sequence update)
DT 01-FEB-1995 (Rel. 31. Last annotation update)
DE HYPOTHETICAL 30.1 KDA PROTEIN IN CONF-TETB INTERGENIC REGION.
GN YBBI.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-168;
RX MEDLINE=96051385; PubMed=7584024;
RA Ogasawara N., Nakai S., Yoshikawa H.;
RT "Systematic sequencing of the 180 kilobase region of the Bacillus

```

```

RT subtilis chromosome containing the replication origin."
RL DNA Res. 1:1-14(1994).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D26185; BAA05194.1; -
DR EMBL: 299124; CAB16100.1; -
DR Subtilisin; Bg10022; YBBI.
KW Hypothetical protein.
SQ SEQUENCE 262 AA; 30149 MW; 1DAAF2F271AB5E4 CRC64;

Query Match 71.4%; Score 30; DB 1; Length 262;
Best Local Similarity 100.0%; Pred. No. 61;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GRLVQ 6
DB 16 GRLVQ 21

RESULT 12
HYBP_ALCEU STANDARD: PRT; 361 AA.
AC P31902;
DT 01-JUL-1993 (Rel. 26. Created)
DT 01-NOV-1995 (Rel. 32. Last sequence update)
DT 15-JUL-1999 (Rel. 38. Last annotation update)
DE HYDROGENASE EXPRESSION/FORMATION PROTEIN HYBP.
GN HYBP.
OS Alkaligenes eutrophus (Ralstonia eutropha).
OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
OC Plasmid megaplasmid pRGL.
OX NCBI_TaxID=510;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-H16 / ATCC 17699;
RX MEDLINE=93356597; PubMed=8352644;
RA Dermede J., Eitinger M., Friedrich B.;
RT "Analysis of a pleiotropic gene region involved in formation of
RT catalytically active hydrogenases in Alkaligenes eutrophus H16."
RL Arch. Microbiol. 159:545-553(1993).
RN [2]
RP REVISIONS.
RC STRAIN-H16 / ATCC 17699;
RA Eitinger T.;
RL Submitted (NOV-1994) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: COULD BE INVOLVED IN NICKEL BINDING AND ACCUMULATION.
CC -1- SIMILARITY: BELONGS TO THE HYBP/HUPM FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X70183; CAA9732.1; -
DR PIR: S29976; S29976.
DR InterPro: IPR002894; -
DR Pfam: PF01495; HYBP_Ureg; 1.
KW Nickel; Plasmid.
SQ SEQUENCE 361 AA; 38483 MW; BC42976B7C5F75C4 CRC64;

```

Query Match 71.4%; Score 30; DB 1; Length 361;  
 Best Local Similarity 75.0%; Pred. No. 84;  
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GRLVODG 8  
 Db 26 GRLVODG 33

RESULT 13  
 ADH3\_ECOLI  
 ID ADH3\_ECOLI STANDARD; PRT; 369 AA.  
 AC P25437; P56596; Q47533;  
 DT 01-MAY-1992 (Rel. 22, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 15-JUL-1998 (Rel. 36, Last annotation update)  
 DE ALCOHOL DEHYDROGENASE CLASS III (EC 1.1.1.1) (GLUTATHIONE-DEPENDENT  
 DE FORMALDEHYDE DEHYDROGENASE) (EC 1.2.1.1) (FDH) (PALDH).  
 GN ADHC.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Escherichia.  
 OX NCBI\_TaxID=562;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-K12;  
 RA Nishimoto H., Saito N.;  
 RA Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Ito K., Matsumoto K., Tsuru D., Yoshimoto T.;  
 RA Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-K12 / MG1655;  
 RA MEDLINE-9742617; PubMed-9278503;  
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.D.,  
 RA Mau B., Shao Y.;  
 RA "The complete genome sequence of Escherichia coli K-12.";  
 RA Science 277:1453-1474(1997).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RA Duncan M., Allen E., Araujo R., Aparicio A.M., Chung E., Davis K.,  
 RA Federgruel N., Hyman R., Kalman S., Komp C., Kurdi O., Lew H., Lin D.,  
 RA Nemeth A., Oefner P., Roberts D., Schramm S., Davis R.W.;  
 RA Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.  
 RN [5]  
 RP SEQUENCE OF 1-47.  
 RX MEDLINE-92118844; PubMed-1731906;  
 RA Guthail W.G., Holmquist B., Vallee B.L.;  
 RT "Purification, characterization, and partial sequence of the  
 RT glutathione-dependent formaldehyde dehydrogenase from Escherichia  
 RT coli: a class III alcohol dehydrogenase.";  
 RL Biochemistry 31:475-481(1992).  
 CC -1- FUNCTION: HAS HIGH FORMALDEHYDE DEHYDROGENASE ACTIVITY IN THE  
 CC PRESENCE OF GLUTATHIONE AND CATALYZES THE OXIDATION OF NORMAL  
 CC ALCOHOLS IN A REACTION THAT IS NOT GSH-DEPENDENT. IN ADDITION,  
 CC HEMIOLACETALS OTHER THAN THOSE FORMED FROM GSH, INCLUDING  
 CC OMEGA-THIOL FATTY ACIDS, ALSO ARE SUBSTRATES.  
 CC -1- CATALYTIC ACTIVITY: ALCOHOL + NAD(+) = ALDEHYDE OR KETONE + NADH.  
 CC -1- CATALYTIC ACTIVITY: FORMALDEHYDE + GLUTATHIONE + NAD(+) =  
 CC S-FORMYLGLUTATHIONE + NADH.  
 CC -1- COFACTOR: REQUIRES ZINC FOR ITS ACTIVITY (4 IONS OF ZINC PER  
 CC DIMER).  
 CC -1- SUBUNIT: HOMODIMER.  
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.  
 CC -1- SIMILARITY: BELONGS TO THE ZINC-CONTAINING ALCOHOL DEHYDROGENASE  
 CC FAMILY. CLASS-III SUBFAMILY.  
 CC -1- CAUTION: REF.1 AND REF.4 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO A  
 CC FRAMESHIFT IN POSITION 26.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
 DR EMBL: D85613; BAA12834.1; ALT\_FRAME.  
 DR EMBL: D38504; BAA22412.1; -.  
 DR EMBL: AEO00142; AAC73459.1; -.  
 DR EMBL: U73857; AAB18081.1; ALT\_FRAME.  
 DR HSP: P11766; 1TEH.  
 DR ECGene; EG50010; adhC.  
 DR InterPro; IPR002085; -.  
 DR InterPro; IPR002328; -.  
 DR Pfam; PF00107; adh\_zinc; 1.  
 DR PROSITE; PS00059; ADH\_ZINC; 1.  
 KW Oxidoreductase; Zinc; NAD.  
 FT METAL 40 40  
 FT METAL 62 62 ZINC (CATALYTIC) (BY SIMILARITY).  
 FT METAL 92 92 ZINC (CATALYTIC) (BY SIMILARITY).  
 FT METAL 95 95 ZINC (SECOND ATOM) (BY SIMILARITY).  
 FT METAL 98 98 ZINC (SECOND ATOM) (BY SIMILARITY).  
 FT METAL 106 106 ZINC (SECOND ATOM) (BY SIMILARITY).  
 FT METAL 169 169 ZINC (CATALYTIC) (BY SIMILARITY).  
 FT CONFLICT 41 41 H -> E (IN REF. 5).  
 FT CONFLICT 46 46 T -> G (IN REF. 5).  
 SQ SEQUENCE 369 AA; 39359 MW; 35B59078F8173521 CRC64;

Query Match 71.4%; Score 30; DB 1; Length 369;  
 Best Local Similarity 62.5%; Pred. No. 86;  
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GRLVODG 8  
 Db 114 GRLVODG 121

RESULT 14  
 ADH3\_HORSE  
 ID ADH3\_HORSE STANDARD; PRT; 373 AA.  
 AC P19854;  
 DT 01-FEB-1991 (Rel. 17, Created)  
 DT 01-FEB-1991 (Rel. 17, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE ALCOHOL DEHYDROGENASE CLASS III CHAIN (EC 1.1.1.1) (GLUTATHIONE-  
 DE DEPENDENT FORMALDEHYDE DEHYDROGENASE) (EC 1.2.1.1) (FDH) (PALDH).  
 GN ADH5.  
 OS Equus caballus (Horse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Perissodactyla; Equidae; Equus.  
 OX NCBI\_TaxID=9796;  
 RN [1]  
 RP SEQUENCE  
 RX MEDLINE-90105360; PubMed-2690942;  
 RA Kaiser R., Holmquist B., Vallee B.L., Joernvall H.;  
 RT "Characteristics of mammalian class III alcohol dehydrogenases, an  
 RT enzyme less variable than the traditional liver enzyme of class I.";  
 RL Biochemistry 28:8432-8438(1989).  
 RN [2]  
 RP SEQUENCE OF 1-6.  
 RX MEDLINE-88005160; PubMed-3653405;  
 RA Fairwell T., Julia P., Kaiser R., Holmquist B., Pares X., Vallee B.L.,  
 RA Joernvall H.;  
 RT "Acetylated N-terminal structures of class III alcohol  
 RT dehydrogenases. Differences among the three enzyme classes.";  
 RL FEBS Lett. 222:99-103(1987).  
 CC -1- FUNCTION: CLASS-III ADH IS REMARKABLY INEFFECTIVE IN OXIDIZING  
 CC ETHANOL, BUT IT READILY CATALYZES THE OXIDATION OF LONG-CHAIN  
 CC PRIMARY ALCOHOLS AND THE OXIDATION OF S-(HYDROXYMETHYL)  
 CC GLUTATHIONE.



CC -1- CATALYTIC ACTIVITY: ALCOHOL + NAD(+) = ALDEHYDE OR KETONE + NADH.  
 CC -1- CATALYTIC ACTIVITY: FORMALDEHYDE + GLUTATHIONE + NAD(+) =  
 CC -1- COFACTOR: REQUIRES ZINC FOR ITS ACTIVITY.  
 CC -1- SUBUNIT: HOMODIMER.  
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.  
 CC -1- SIMILARITY: BELONGS TO THE ZINC-CONTAINING ALCOHOL DEHYDROGENASE  
 DR PIR: S02617, S02617.  
 DR PIR: A33419, A33419.  
 DR HSSP: P11766, I1766.  
 DR InterPro: IPR002085.  
 DR Pfam: PF00107, adh\_zinc.1.  
 DR PROSITE: PS00059; ADH\_ZINC.1.  
 KW Oxidoreductase; Zinc; NAD; Multigene family; Acetylation.  
 FT MOD\_RES 1 1 ACETYLATION (BY SIMILARITY).  
 FT METAL 44 44 ZINC (CATALYTIC).  
 FT METAL 66 66 ZINC (CATALYTIC).  
 FT METAL 96 96 ZINC (CATALYTIC).  
 FT METAL 99 99 ZINC (SECOND ATOM).  
 FT METAL 102 102 ZINC (SECOND ATOM).  
 FT METAL 110 110 ZINC (SECOND ATOM).  
 FT METAL 173 173 ZINC (SECOND ATOM).  
 FT BINDING 114 114 ZINC (CATALYTIC).  
 FT BY FATY ACIDS (BY SIMILARITY).  
 SQ SEQUENCE 373 AA; 39440 MW; 491F01B0A7D43D32 CRC64;  
 Query Match 71.4%; Score 30; DB 1; Length 373;  
 Best Local Similarity 62.5%; Pred. No. 87;  
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 GRLVYDGG 8  
 DB 118 GKGLMPDG 125  
 RESULT 15  
 ADHX\_HUMAN STANDARD; PRT; 373 AA.  
 AC P11766;  
 DT 01-OCT-1989 (Rel. 12, Created)  
 DT 01-APR-1990 (Rel. 14, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE ALCOHOL DEHYDROGENASE CLASS III CHT CHAIN (EC 1.1.1.1) (GLUTATHIONE-  
 DE DEPENDENT FORMALDEHYDE DEHYDROGENASE) (EC 1.2.1.1) (FDH).  
 GN ADHS OR ADHX OR FDH.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OC NCBI\_TaxID:9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-90056459; PubMed-2818582;  
 RA Sharma C.P., Fox E.A., Holmquist B., Joernvall H., Vallee B.L.;  
 RT "CDNA sequence of human class III alcohol dehydrogenase.";  
 RL Biochem. Biophys. Res. Commun. 164:631-637(1989).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-90026418; PubMed-2679557;  
 RA Gira P.R., King J.F., Kozak C., Moretti T., O'Brien S.J.,  
 RT "Cloning and comparative mapping of a human class III (chl) alcohol  
 RL dehydrogenase cDNA.";  
 RL Biochem. Biophys. Res. Commun. 164:453-460(1989).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-93077045; PubMed-1446828;  
 RA Hur M.W., Edenberg H.J.;  
 RT "Cloning and characterization of the ADHS gene encoding human alcohol  
 RL dehydrogenase 5, formaldehyde dehydrogenase.";  
 RL Gene 121:305-311(1992).

RN [4]  
 RP SEQUENCE.  
 RC TISSUE-Liver;  
 RX MEDLINE-88209465; PubMed-3365377;  
 RA Kaiser R., Holmquist B., Hempel J., Vallee B.L., Joernvall H.;  
 RT "Class III human liver alcohol dehydrogenase: a novel structural type  
 RL Biochemistry 27:1132-1140(1988).  
 RN [5]  
 RP PARTIAL SEQUENCE, AND MUTAGENESIS OF ARG-114.  
 RX MEDLINE-93264427; PubMed-8494891;  
 RA Holmquist B., Moulis J.M., Engeland K., Vallee B.L.;  
 RT "Role of arginine 115 in fatty acid activation and formaldehyde  
 RL dehydrogenase activity of human class III alcohol dehydrogenase.";  
 RN Biochemistry 32:5139-5144(1993).  
 RP MUTAGENESIS OF ARG-114.  
 RX MEDLINE-93211987; PubMed-8460164;  
 RA Engeland K., Hoeoeg J.-O., Holmquist B., Estenius M., Joernvall H.,  
 RA Vallee B.L.;  
 RT "Mutation of Arg-115 of human class III alcohol dehydrogenase: a  
 RL binding site required for formaldehyde dehydrogenase activity and  
 RT fatty acid activation.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 90:2491-2494(1993).  
 RN [7]  
 RP X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS).  
 RX MEDLINE-97170743; PubMed-9018047;  
 RA Yang Z.-N., Boston W.P., Hurley T.D.;  
 RT "Structure of human chl chl alcohol dehydrogenase: a glutathione-  
 RL dependent formaldehyde dehydrogenase.";  
 CC J. Mol. Biol. 265:330-343(1997).  
 CC -1- FUNCTION: CLASS-III ADH IS REMARKABLY INEFFECTIVE IN OXIDIZING  
 CC ETHANOL, BUT IT READILY CATALYZES THE OXIDATION OF LONG-CHAIN  
 CC PRIMARY ALCOHOLS AND THE OXIDATION OF S-(HYDROXYMETHYL)  
 CC GLUTATHIONE.  
 CC -1- CATALYTIC ACTIVITY: ALCOHOL + NAD(+) = ALDEHYDE OR KETONE + NADH.  
 CC -1- CATALYTIC ACTIVITY: FORMALDEHYDE + GLUTATHIONE + NAD(+) =  
 CC S-FORMYLGLUTATHIONE + NADH.  
 CC -1- COFACTOR: REQUIRES ZINC FOR ITS ACTIVITY.  
 CC -1- SUBUNIT: HOMODIMER.  
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.  
 CC MISCELLANEOUS: THERE ARE 7 DIFFERENT ADH'S ISOZYMES IN HUMAN:  
 CC PI, ONE TO CLASS-I: ALPHA, BETA, AND GAMMA, ONE TO CLASS-II:  
 CC CLASS-V: ADH6.  
 CC -1- SIMILARITY: BELONGS TO THE ZINC-CONTAINING ALCOHOL DEHYDROGENASE  
 CC FAMILY. CLASS-III SUBFAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC use by non-profit institutions. There are no restrictions on its  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC or send an email to license@lsb.slb.ch).  
 DR EMBL: M30471; AAA79018.1;  
 DR EMBL: M29872; AAA51597.1;  
 DR EMBL: M81118; AAA51596.1;  
 DR EMBL: M81113; AAA51596.1; JOINED.  
 DR EMBL: M81114; AAA51596.1; JOINED.  
 DR EMBL: M81115; AAA51596.1; JOINED.  
 DR EMBL: M81116; AAA51596.1; JOINED.  
 DR EMBL: M81117; AAA51596.1; JOINED.  
 DR PIR: A33428; DEHNC2.  
 DR PIR: A36739; A36739.  
 DR PIR: JH0789; JH0789.  
 DR PDB: 1TBH; 07-DEC-96.  
 DR MIM: 103710;  
 DR MIM: 136490;  
 DR InterPro: IPR002085;  
 DR InterPro: IPR002328;



DR Pfam: PF00107: adh\_zinc: 1.  
DR PROSITE: PS00059: ADH\_ZINC: 1.  
KW Oxidoreductase; Zinc; NAD; Multigene family; Acetylation;  
3D-structure.  
FT INIT\_MET 0 0  
FT MOD\_RES 1 1 ACETYLATION.  
FT METAL 44 44 ZINC (CATALYTIC).  
FT METAL 66 66 ZINC (CATALYTIC).  
FT METAL 96 66 ZINC (SECOND ATOM).  
FT METAL 99 99 ZINC (SECOND ATOM).  
FT METAL 102 102 ZINC (SECOND ATOM).  
FT METAL 110 110 ZINC (SECOND ATOM).  
FT METAL 173 173 ZINC (CATALYTIC).  
FT BINDING 114 114 IMPORTANT FOR FDH ACTIVITY AND ACTIVATION  
BY FATTY ACIDS.  
FT MUTAGEN 114 114 R->A,D: LOSS OF FDH ACTIVITY, AND LOSS OF  
ACTIVATION BY FATTY ACIDS.  
FT CONFLICT 166 166 D -> Y (IN REF. 2).  
FT CONFLICT 245 245 F -> L (IN REF. 2).  
SQ SEQUENCE 373 AA; 39593 MW; A4E1A8D6F0424F4 CRC64;

Query Match 71.4%; Score 30; DB 1; Length 373;  
Best Local Similarity 62.5%; Pred. No. 87;  
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 GRGLVDDG 8  
1:11:11  
DB 118 GKGLMPDG 125

Search completed: June 13, 2001, 14:21:45  
Job time: 805 sec

*This Page Blank (uspto)*

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 13, 2001, 14:20:31 ; Search time 225.85 Seconds  
(without alignments)  
4.152 Million cell updates/sec

Title: PCT-US01-05825A-12

Perfect score: 42

Sequence: 1 GRGLVDDG 8

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 374700 seqs, 117207915 residues

Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL.15:\*

1: sp.archaea:\*\n2: sp.bacteria:\*\n3: sp.fungi:\*\n4: sp.human:\*\n5: sp.invertebrate:\*\n6: sp.mammal:\*\n7: sp.mhc:\*\n8: sp.organelle:\*\n9: sp.phage:\*\n10: sp.plant:\*\n11: sp rodent:\*\n12: sp.unclassified:\*\n13: sp.vertebrate:\*\n14: sp.virus:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	37	88.1	540	4	Q9Y3Y3 homo sapien
2	37	88.1	671	4	Q9NRF2
3	37	88.1	683	4	Q9NRF1
4	37	88.1	730	4	Q9P2P7
5	37	88.1	756	4	Q9NRF3
6	37	88.1	982	2	Q33492
7	35	83.3	288	2	Q9P366
8	35	83.3	580	2	Q35856
9	35	83.3	942	3	Q9UV85
10	35	83.3	942	3	Q9UV84
11	34	81.0	152	10	Q9L157
12	34	81.0	687	2	Q9KFF3
13	34	81.0	1353	13	Q00546
14	33	78.6	207	2	Q52737
15	33	78.6	295	2	Q9K9C9
16	33	78.6	378	3	Q9UUF6
17	33	78.6	464	2	Q87268
18	33	78.6	482	1	Q29120
19	33	78.6	609	10	Q9LR69

20	33	78.6	3436	14	Q66666
21	32	76.2	264	5	Q9NH66
22	32	76.2	281	1	Q9YC28
23	32	76.2	350	2	Q43986
24	32	76.2	686	5	Q94316
25	32	76.2	724	10	Q9L1E4
26	32	76.2	739	5	Q9N8U6
27	32	76.2	780	10	Q9SAJ5
28	32	76.2	925	10	Q92TA5
29	32	76.2	940	3	Q9UV87
30	32	76.2	1158	3	Q9UTR5
31	32	76.2	1194	10	Q9SHF2
32	32	76.2	1236	2	Q06264
33	31	73.8	57	14	Q96595
34	31	73.8	117	14	Q95387
35	31	73.8	122	14	Q9YR84
36	31	73.8	163	10	P93330
37	31	73.8	199	2	Q9ZB14
38	31	73.8	245	5	Q9YVB8
39	31	73.8	278	2	Q9KVT3
40	31	73.8	309	5	Q9VNC5
41	31	73.8	448	5	Q61540
42	31	73.8	448	5	Q9VLR9
43	31	73.8	449	14	P88946
44	31	73.8	526	2	Q9K1X5
45	31	73.8	526	2	Q9JS76

## ALIGNMENTS

RESULT	ID	Q9Y3Y3	PRELIMINARY:	PRT:	540 AA.
AC	Q9Y3Y3				
DT	01-NOV-1999	(TREMBLrel. 12, Created)			
DT	01-NOV-1999	(TREMBLrel. 12, Last sequence update)			
DT	01-OCT-2000	(TREMBLrel. 15, Last annotation update)			
DE	HYPOTHETICAL 57.3 KDA PROTEIN (FRAGMENT).				
GN	DKF2P54/G1110.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
OX	NCBI_TaxID=9606;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=BRAIN;				
RA	Poustka A., Klein M., Mewes H.W., Gassenhuber J., Wlmann S.;				
RL	Submitted (MAY-1999) to the EMBL/Genbank/DBJ databases.				
DR	EMBL; AL04924; CAB43208.1; -.				
DR	HSSP; P29353; IMIL.				
DR	INTERPRO; IPR000980; -.				
DR	INTERPRO; IPR001849; -.				
DR	PFAM; PF00017; SH2.1.				
DR	PFAM; PF00169; PH.1.				
DR	PRINTS; PR00401; SHDOMAIN.				
DR	PROSITE; PS50001; SH2.1.				
KW	HYPOTHETICAL protein.				
FT	NON_TER				
SO	SEQUENCE	540 AA; 57334 MW; EB9806C8CB5EBD3 CRC64;			

Query Match 88.1%; Score 37; DB 4; Length 540;  
Best Local Similarity 87.5%; Pred. No. 22;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY	1	GRGLVDDG 8
DB	67	GRGLVSDG 74

RESULT 2

...ata; Euteleostomi;  
...; Homnidae; Homo.

RESULT	6	
ID	033492	
AC	033492	
DT	01-JAN-1998	PRELIMINARY;
DT	01-JAN-1998	(T)EMBLrel. 05, PRT;
DT	01-OCT-2000	(T)EMBLrel. 05, Created
DE	STYRENE SENSOR KINASE.	15, last seq
OS	Pseudomonas sp.	last annotation update)
OC	Bacteriis; Proteobacteria.	
NCBI_TaxID=306;		
SEQUENCE FROM N.A.		
MEDLINE=96155131; Pubmed=9495743;		
Genetic and Alonso S. Garcia J.L.,		
Pseudomonas sp. strain analysis of the		
Bacteriis sp. strain analysis of the		
EMB1; AF000330; CA03998 I, (1998)		

DR INTERPRO: IPR000014; -  
 DR INTERPRO: IPR000010; -  
 DR INTERPRO: IPR000700; -  
 DR INTERPRO: IPR001610; -  
 DR INTERPRO: IPR001789; -  
 DR PFAM: PF00072; response\_reg; 1.  
 DR PFAM: PF00512; signal; 2.  
 DR PFAM: PF00785; PAC; 1.  
 DR PFAM: PF00989; PAS; 1.  
 DR PRINTS: PR00344; BCTRLSENSOR.  
 DR SEQUENCE 982 AA; 108745 MW; ED209EADDCE579AD CRC64;

Query Match 88.1%; Score 37; DB 2; Length 982;  
 Best Local Similarity 87.5%; Pred. No. 43;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GRLVODG 8  
 |||||  
 DB 439 GRLVSDG 446

RESULT 7  
 O9P366 PRELIMINARY; PRT: 288 AA.  
 AC O9P366.  
 DT 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)  
 DE CONSERVED HYPOTHETICAL PROTEIN FRAGMENT.  
 GN 13E11.10.  
 OS Neurospora crassa.  
 OC Eukaryota; fungi; Ascomycota; Sordariatales; Sordariaceae; Neurospora.  
 OX NCBI\_TaxID=5141;  
 RN [1]  
 RA SEQUENCE FROM N.A.  
 RA Schulte U., Aign V., Hohnsels J., Brandt P., Fartmann B., Holland R.,  
 RA Nykterura G., Meves H.W., Manhaupt G.;  
 RL Submitted (Apr-2000) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RA SEQUENCE FROM N.A.  
 RA German Neurospora genome project;  
 RL Submitted (Apr-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AL353820; CAB8569.1; -  
 SQ SEQUENCE 288 AA; 32004 MW; AA47812F595AB1FD CRC64;

Query Match 83.3%; Score 35; DB 3; Length 288;  
 Best Local Similarity 87.5%; Pred. No. 28;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GRLVODG 8  
 |||||  
 DB 225 GRLVVDG 232

RESULT 8  
 O55856 PRELIMINARY; PRT: 580 AA.  
 AC O55856.  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
 DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)  
 DE ABC TRANSPORTER.  
 GN SLR0615.  
 OS Synechocystis sp. (strain PCC 6803).  
 OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.  
 OX NCBI\_TaxID=1148;  
 RN [1]  
 RA SEQUENCE FROM N.A.  
 RP STRAIN-PCC6803;  
 RX MEDLINE=96127529; PubMed=8590279;  
 RA Kaneko T., Tanaka A., Sato S., Kotani H., Sazuka T., Miyajima N.,

RA Sugita M., Tabata S.;  
 RT "Sequence analysis of the genome of the unicellular cyanobacterium  
 RT Synechocystis sp. strain PCC6803. I. Sequence features in the 1 Mb  
 RT region from map positions 64% to 92% of the genome.";  
 RL DNA Res. 2:153-166(1995).  
 RN [2]

RP SEQUENCE FROM N.A.  
 RX MEDLINE=97061201; PubMed=8905231;  
 RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,  
 RA Miyajima N., Hikosawa M., Sugita M., Sasamoto S., Kimura T.,  
 RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S.,  
 RA Shimo S., Takeuchi C., Wada T., Watanabe A., Yamada M.,  
 RA Tabata S.;  
 RT "Sequence analysis of the genome of the unicellular cyanobacterium  
 RT Synechocystis sp. strain PCC6803. II. Sequence determination of the  
 RT entire genome and assignment of potential protein-coding regions.";  
 RL DNA Res. 3:109-136(1996).  
 DR EMBL: D64004; BAA10612.1; -  
 DR INTERPRO: IPR000595; -  
 DR INTERPRO: IPR001140; -  
 DR INTERPRO: IPR001617; -  
 DR PFAM: PF000005; ABC\_tran; 1.  
 DR PFAM: PF00664; ABC\_membrane; 1.  
 DR PROSITE: PS00211; ABC\_TRANSPORTER; UNKNOWN\_1.  
 DR PROSITE: PS00889; CNMP\_BINDING\_2; UNKNOWN\_1.  
 SQ SEQUENCE 580 AA; 64292 MW; 3D5ADD81B67D7C1 CRC64;

Query Match 83.3%; Score 35; DB 2; Length 580;  
 Best Local Similarity 75.0%; Pred. No. 61;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 GRLVODG 8  
 |||||  
 DB 263 GTLIODG 270

RESULT 9  
 O9UV85 PRELIMINARY; PRT: 942 AA.  
 AC O9UV85.  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)  
 DE DNA-DEPENDENT RNA POLYMERASE II RPB140 (FRAGMENT).  
 GN RPB2.  
 OS Curvularia brachyspora.  
 OC Eukaryota; Fungi; Ascomycota; Pleosporales; Pleosporaceae;  
 OC anamorphic Pleosporaceae; Curvularia.  
 OX NCBI\_TaxID=39400;  
 RN [1]  
 RN RP SEQUENCE FROM N.A.  
 RC STRAIN-ATCC12330;  
 RA Liu Y.J., Whelen S., Hall B.D.;  
 RT "Phylogenetic relationships among ascomycetes: evidence from an RNA  
 RT polymerase II subunit.";  
 RL Mol. Biol. Evol. 16:1799-1808(1999).  
 DR EMBL: AF107803; AAF19075.1; -  
 DR INTERPRO: IPR001572; -  
 DR PFAM: PF00562; RNA\_pol\_B; 2.  
 DR PROSITE: PS01166; RNA\_POL\_BETA; 1.  
 FT NON\_TER 1  
 FT NON\_TER 942  
 FT NON\_TER 942  
 SQ SEQUENCE 942 AA; 106398 MW; FB508BDD91A90982 CRC64;

Query Match 83.3%; Score 35; DB 3; Length 942;  
 Best Local Similarity 85.7%; Pred. No. 11e+02;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 RGLVODG 8  
 |||||  
 DB 485 RGLIODG 491

RESULT 10  
 ID Q9UV84 PRELIMINARY; PRT; 942 AA.  
 AC Q9UV84;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DE 01-JUN-2000 (TREMBLrel. 14, Last annotation update)  
 GN DNA-DEPENDENT RNA POLYMERASE II RPB140 (FRAGMENT).  
 OS Pleospora herbarum.  
 OC Eukaryota; Fungi; Ascomycota; Pleosporales; Pleosporaceae; Pleospora.  
 OX NCBI\_TaxID=31177;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=EGS04-188C;  
 RA Liu Y.J., Whelen S., Hall B.D.;  
 RT "Phylogenetic relationships among ascomycetes: evidence from an RNA  
 polymerase II subunit."  
 RL Mol. Biol. Evol. 16:1799-1808(1999).  
 DR EMBL; AF107804; AAF19076.1;  
 DR INTERPRO: IPR001572;  
 DR PFAM: PF00562; RNA\_POL\_B; 1.  
 DR PROSITE: PS01166; RNA\_POL\_BETA; 1.  
 FT NON\_TER  
 FT NON\_TER  
 SQ SEQUENCE 942 AA; 106436 MW; E61510108C771078 CRC64;

Query Match  
 Best Local Similarity 83.3%; Score 35; DB 3; Length 942;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 GRLVODG 8  
 DB 485 GRLVODG 491

RESULT 11  
 ID Q9L157 PRELIMINARY; PRT; 152 AA.  
 AC Q9L157;  
 DT 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DE 01-OCT-2000 (TREMBLrel. 15, Last annotation update)  
 GN Oryza sativa (Rice).  
 OS Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 OC Magnoliophyta; Liliopsida; Poales; Poaceae; Oryza.  
 OX NCBI\_TaxID=4530;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. NIPPONBARE;  
 RA Sasaki T., Matsumoto T., Yamamoto K.;  
 RT "Oryza sativa nipponebare(GA3) genomic DNA, chromosome 1, PAC  
 clone:PO469E09."  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AP001366; BA692410.1;  
 SQ SEQUENCE 152 AA; 16852 MW; 5D32B935DE81D619 CRC64;

Query Match  
 Best Local Similarity 81.0%; Score 34; DB 10; Length 152;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 GRLVODG 8  
 DB 24 GRLVODG 31

ID Q9KYF3 PRELIMINARY; PRT; 687 AA.  
 AC Q9KYF3;  
 DT 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DE 01-OCT-2000 (TREMBLrel. 15, Last annotation update)  
 GN PUTATIVE MEMBRANE PROTEIN.  
 OS Streptomyces coelicolor.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.  
 OX NCBI\_TaxID=1902;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A3(2);  
 RA Brown S.P., Harris D.;  
 RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A3(2);  
 RA Bentley S.D., Parkhill J., Barrell B.G., Rajandream M.A.;  
 RN Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A3(2);  
 RA Redenbach M., Kieser H.M., Denapalte D., Eichner A., Cullum J.,  
 RA Kinashi H., Hopwood D.A.;  
 RT "A set of ordered cosmids and a detailed genetic and physical map for  
 the 8 Mb Streptomyces coelicolor A3(2) chromosome."  
 RL Mol. Microbiol. 21:77-96(1996).  
 DR EMBL; AL356612; CAB92370.1;  
 SQ SEQUENCE 687 AA; 70938 MW; B6AD2F58253C5889 CRC64;

Query Match  
 Best Local Similarity 81.0%; Score 34; DB 2; Length 687;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 GRLVODG 8  
 DB 63 GRLVODG 70

RESULT 13  
 ID Q00546 PRELIMINARY; PRT; 1353 AA.  
 AC Q00546;  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
 DE 01-JUN-2000 (TREMBLrel. 14, Last annotation update)  
 GN Gallus gallus (Chicken).  
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 OX NCBI\_TaxID=9031;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=BRAIN;  
 RX MEDLINE=92265298; PubMed=1375037;  
 RA Noerenberg U., Wille H., Wolff M., Frank R., Rathjen F.G.;  
 RT "The chicken neural extracellular matrix molecule restrictin:  
 RT motifs with EGF-, fibronectin type III-, and fibrinogen-like  
 motifs."  
 RL Neuron 8:849-863(1992)  
 DR EMBL; X64649; CAA45920.1;  
 DR PIR: JH0675; JH0675.  
 DR HSSP: P24821; ITEN.  
 DR INTERPRO: IPR000561;  
 DR INTERPRO: IPR000950;  
 DR INTERPRO: IPR001777;  
 DR INTERPRO: IPR002181;  
 DR PFAM: PF00008; EGF; 4.  
 DR PFAM: PF00041; fn3; 9.

DR PFAM: PF00147; fibrinogen.C; 1.  
 DR PROSITE: PS00022; EGF\_1; UNKNOWN.5.  
 DR PROSITE: PS00241; RECEPTOR\_CYTOKINES\_1; UNKNOWN.1.  
 DR PROSITE: PS01186; EGF\_2; 4.  
 KW SIGNAL.  
 FT SIGNAL 1 33 POTENTIAL.  
 FT CHAIN 34 1353 RESTRICTIN.  
 SQ SEQUENCE 1353 AA; 148278 MW; CD8393C4203171D9 CRC64;

Query Match 81.0%; Score 34; DB 13; Length 1353;  
 Best Local Similarity 75.0%; Pred. No. 2.5e+02;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 GRGLVODG 8  
 |||: |||  
 DB 302 GRGVCDG 309

RESULT 14  
 052737  
 ID 052737 PRELIMINARY; PRT; 207 AA.  
 AC 052737;  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
 DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)  
 DE NITRILE HYDRATASE ALPHA SUBUNIT.  
 OS Rhodococcus erythropolis.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Corynebacterineae; Nocardiaceae; Rhodococcus.  
 OX NCBI\_TaxID=1833;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-JCM6823;  
 RX MEDLINE=93379358; PubMed=7764017;  
 RA Robert D., Nishiyama M., Horiouchi S., Beppu T.;  
 RT "Characterization of nitrile hydratase genes cloned by DNA screening  
 from Rhodococcus erythropolis";  
 RL Biosci. Biotechnol. Biochem. 57:1323-1328(1993).  
 DR EMBL: D14454; BAA03348.1; -.  
 DR HSSP: P13448; IAHJ.  
 SQ SEQUENCE 207 AA; 23116 MW; F331BC6A11FCA394 CRC64;

Query Match 78.6%; Score 33; DB 2; Length 207;  
 Best Local Similarity 75.0%; Pred. No. 49;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 GRGLVODG 8  
 |||: |||  
 DB 32 GKGLVODG 39

RESULT 15  
 09K9C9  
 ID 09K9C9 PRELIMINARY; PRT; 295 AA.  
 AC 09K9C9;  
 DT 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)  
 DE BH2720 PROTEIN.  
 GN BH2720.  
 OS Bacillus halodurans.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
 OC Bacillus/Staphylococcus group; Bacillus.  
 OX NCBI\_TaxID=86665;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-C-125 / JCM 9153;  
 RA Takami H., Nakasone K., Takaki Y.;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AP001516; BAB06439.1; -.  
 SQ SEQUENCE 295 AA; 34308 MW; 1BDE317EDA39B764 CRC64;

Query Match 78.6%; Score 33; DB 2; Length 295;  
 Best Local Similarity 87.5%; Pred. No. 74;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GRGLVODG 8  
 |||||: |  
 DB 220 GRGLVODG 227

Search completed: June 13, 2001, 14:20:32  
 Job time: 732 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OW protein - protein search, using sw model

Run on: June 13, 2001, 14:16:34 ; Search time 118.55 Seconds  
(without alignments)  
1.296 Million cell updates/sec

Title: PCT-US01-05825A-12

Perfect score: 42

Sequence: 1 GRLVODG 8

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 185757 seqs, 19210857 residues

Total number of hits satisfying chosen parameters: 185757

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents: AA:\*  
1: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep:\*  
2: /cgn2\_6/ptodata/2/1aa/5B.COMB.pep:\*  
3: /cgn2\_6/ptodata/2/1aa/6A.COMB.pep:\*  
4: /cgn2\_6/ptodata/2/1aa/6B.COMB.pep:\*  
5: /cgn2\_6/ptodata/2/1aa/PCITUS.COMB.pep:\*  
6: /cgn2\_6/ptodata/2/1aa/Backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	35	83.3	246	1	US-07-843-125-11 Sequence 11, Appl
2	31	73.8	117	4	US-08-973-068-5 Sequence 5, Appl
3	31	73.8	963	1	US-08-537-002A-3 Sequence 3, Appl
4	31	73.8	963	3	US-08-863-010-3 Sequence 3, Appl
5	31	73.8	963	4	US-09-024-428-3 Sequence 3, Appl
6	30	71.4	90	1	US-08-464-339A-12 Sequence 12, Appl
7	30	71.4	316	3	US-08-468-847B-12 Sequence 14, App
8	30	71.4	374	1	US-08-459-101A-2 Sequence 12, Appl
9	30	71.4	375	2	US-08-468-847B-11 Sequence 11, Appl
10	30	71.4	810	2	US-08-820-170A-34 Sequence 34, Appl
11	30	71.4	810	4	US-09-055-699-34 Sequence 34, Appl
12	30	71.4	810	4	US-09-273-565-34 Sequence 34, Appl
13	29	69.0	118	2	US-08-652-816A-12 Sequence 12, Appl
14	29	69.0	119	2	US-08-318-157B-2 Sequence 2, Appl
15	29	69.0	376	1	US-08-608-241-2 Sequence 2, Appl
16	29	69.0	376	2	US-08-922-182-2 Sequence 2, Appl
17	29	69.0	376	2	US-08-919-953-2 Sequence 2, Appl
18	29	69.0	2154	4	US-08-841-349-4 Sequence 4, Appl
19	29	69.0	2639	4	US-09-080-983-3 Sequence 3, Appl
20	29	69.0	3739	4	US-09-320-878-2 Sequence 2, Appl
21	28	66.7	15	1	US-08-331-398A-66 Sequence 66, Appl
22	28	66.7	15	2	US-08-652-558-42 Sequence 66, Appl
23	28	66.7	15	2	US-08-331-397B-66 Sequence 66, Appl
24	28	66.7	15	2	US-08-759-804A-65 Sequence 65, Appl
25	28	66.7	18	1	US-08-331-398A-55 Sequence 55, Appl
26	28	66.7	18	1	US-08-401-908-14 Sequence 14, Appl
27	28	66.7	18	1	US-08-401-908-14 Sequence 14, Appl

28	28	66.7	18	2	US-08-331-397B-55	Sequence 55, Appl
29	28	66.7	18	2	US-08-759-804A-54	Sequence 54, Appl
30	28	66.7	20	2	US-08-859-931A-2	Sequence 2, Appl
31	28	66.7	20	2	US-08-053-451B-113	Sequence 113, App
32	28	66.7	26	1	US-08-471-780C-80	Sequence 80, Appl
33	28	66.7	26	1	US-08-467-282B-80	Sequence 80, Appl
34	28	66.7	26	2	US-08-471-282A-80	Sequence 80, Appl
35	28	66.7	26	2	US-08-466-710C-80	Sequence 80, Appl
36	28	66.7	26	3	US-08-468-739C-80	Sequence 80, Appl
37	28	66.7	29	1	US-08-471-780C-119	Sequence 119, App
38	28	66.7	29	1	US-08-467-282B-119	Sequence 119, App
39	28	66.7	29	2	US-08-471-282A-119	Sequence 119, App
40	28	66.7	29	2	US-08-466-710C-119	Sequence 119, App
41	28	66.7	29	3	US-08-468-739C-119	Sequence 119, App
42	28	66.7	30	1	US-07-988-925-7	Sequence 7, Appl
43	28	66.7	30	1	US-07-977-696C-75	Sequence 75, Appl
44	28	66.7	30	1	US-08-129-930B-75	Sequence 75, Appl
45	28	66.7	30	2	US-08-362-780-7	Sequence 7, Appl

## ALIGNMENTS

RESULT 1  
US-07-843-125-11  
Sequence 11, Application US/07843125  
Patent No. 5395750  
GENERAL INFORMATION:  
APPLICANT: Dillon, Patrick J  
TITLE OF INVENTION: Methods for Producing Proteins which  
TITLE OF INVENTION: Bind to Predetermined Antigens  
NUMBER OF SEQUENCES: 13  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: George M. Gould, Esq., Hoffmann-La Roche Inc.  
STREET: 340 Kingsland Street  
CITY: Nutley  
STATE: New Jersey  
COUNTRY: U.S.A.  
ZIP: 07110-1199  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/843,125  
FILING DATE: 19920228  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Roseman, Catherine R  
REGISTRATION NUMBER: 34240  
REFERENCE/DOCKET NUMBER: 8594  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (201) 235-6208  
TELEFAX: (201) 235-3500  
INFORMATION FOR SEQ ID NO: 11:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 246 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-07-843-125-11

Query Match 83.3%; Score 35; DB 1; Length 246;  
Best Local Similarity 87.5%; Pred. No. 14;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GRLVODG 8  
|||||  
Db 8 GRLVODG 15

```
RESULT 2
US-08-973-068-5
; Sequence 5, Application US/08973068
; Patent No. 6127604
; GENERAL INFORMATION:
; APPLICANT: Dale, James Langham
; APPLICANT: Harding, Robert Maxwell
; APPLICANT: Dugdale, Benjamin
; APPLICANT: Beetham, Peter Ronald
; APPLICANT: Hafner, Gregory John
; TITLE OF INVENTION: INTERGENIC REGIONS OF BANANA BUNCHY TOP VIRUS
; FILE REFERENCE: 09657/002001
; CURRENT APPLICATION NUMBER: US/08/973,068
; EARLIER FILING DATE: 1998-03-12
; EARLIER APPLICATION NUMBER: PCT/AU96/00335
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: FASTSEQ for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 117
; TYPE: PRT
; ORGANISM: Banana Bunchy Top Virus (BBTV)
US-08-973-068-5

Query Match
Best Local Similarity 73.8%; Score 31; DB 4; Length 117;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 GRLVOD 7
Db 75 GRLVED 81

RESULT 3
US-08-537-002A-3
; Sequence 3, Application US/08537002A
; Patent No. 5773282
; GENERAL INFORMATION:
; APPLICANT: TSUSAKI, Keiji
; APPLICANT: KUBOTA, Michio
; APPLICANT: SUGIMOTO, Toshiyuki
; TITLE OF INVENTION: RECOMBINANT THERMOSTABLE ENZYME FOR
; CONVERTING MALTOSE INTO TREHALOSE
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEIMARK
; STREET: 419 Seventh Street, N.W., Suite 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/537,002A
; FILING DATE: 29-SEP-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 260984/1994
; FILING DATE: 01-OCT-1994
; APPLICATION NUMBER: JP No. 5773282 yet received
; ATTORNEY/AGENT INFORMATION:
; NAME: BROWDY, Roger L.
; REGISTRATION NUMBER: 25,618
; REFERENCE/DOCKET NUMBER: TSUSAKI-1
```

```
TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
; TELEX: 248633
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 963 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-537-002A-3

Query Match
Best Local Similarity 73.8%; Score 31; DB 1; Length 963;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 GLVODG 8
Db 725 GLVODG 730

RESULT 4
US-08-863-010-3
; Sequence 3, Application US/08863010
; Patent No. 6087146
; GENERAL INFORMATION:
; APPLICANT: TSUSAKI, Keiji
; APPLICANT: KUBOTA, Michio
; APPLICANT: SUGIMOTO, Toshiyuki
; TITLE OF INVENTION: RECOMBINANT THERMOSTABLE ENZYME FOR
; CONVERTING MALTOSE INTO TREHALOSE
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEIMARK
; STREET: 419 Seventh Street, N.W., Suite 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/863,010
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/537,002
; FILING DATE: 29-SEP-1995
; APPLICATION NUMBER: JP 260984/1994
; FILING DATE: 01-OCT-1994
; APPLICATION NUMBER: JP No. 6087146 yet received
; ATTORNEY/AGENT INFORMATION:
; NAME: BROWDY, Roger L.
; REGISTRATION NUMBER: 25,618
; REFERENCE/DOCKET NUMBER: TSUSAKI-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
; TELEX: 248633
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 963 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
```

US-08-863-010-3

Query Match 73.8%; Score 31; DB 3; Length 963;  
Best Local Similarity 100.0%; Pred. No. 3.4e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 GIVODG 8  
DB 725 GIVODG 730

RESULT 5  
US-09-024-429-3  
; Sequence 3, Application US/09024429  
; Patent No. 6165768  
; GENERAL INFORMATION:  
; APPLICANT: TSUSAKI, Keiji  
; APPLICANT: KUBOTA, Michio  
; APPLICANT: SUGIMOTO, Toshiyuki  
; TITLE OF INVENTION: RECOMBINANT THERMOSTABLE ENZYME FOR  
; TITLE OF INVENTION: CONVERTING MALTOSE INTO TREHALOSE  
; NUMBER OF SEQUENCES: 17  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BROWDY AND NEIMARK  
; STREET: 419 Seventh Street, N.W., Suite 300  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20004  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/024.429  
; FILING DATE:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/537,002  
; FILING DATE: 29-SEP-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 260984/1994  
; FILING DATE: 01-OCT-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 255829/1995  
; FILING DATE: 08-SEP-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: YUN, Allen C.  
; REGISTRATION NUMBER: 37,971  
; REFERENCE/DOCKET NUMBER: TSUSAKI-1B  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-628-5197  
; TELEFAX: 202-737-3528  
; TELEX: 248633  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 963 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; US-09-024-429-3

Query Match 73.8%; Score 31; DB 4; Length 963;  
Best Local Similarity 100.0%; Pred. No. 3.4e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 3 GIVODG 8  
DB 725 GIVODG 730

RESULT 6  
US-08-464-339A-12

; Sequence 12, Application US/08464339A  
; Patent No. 5747280  
; GENERAL INFORMATION:  
; APPLICANT: HASTINGS, ET AL.  
; TITLE OF INVENTION: Human Vascular IBP-Like Growth  
; TITLE OF INVENTION: Factor  
; NUMBER OF SEQUENCES: 17  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,  
; ADDRESSEE: CECCHI, STEWART & OLSTEIN  
; STREET: 6 BECKER FARM ROAD  
; CITY: ROSELAND  
; STATE: NEW JERSEY  
; COUNTRY: USA  
; ZIP: 07068  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5 INCH DISKETTE  
; COMPUTER: IBM PS/2  
; OPERATING SYSTEM: MS-DOS  
; SOFTWARE: WORD PERFECT 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/464,339A  
; FILING DATE: June 5, 1995  
; CLASSIFICATION: 536  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US94/14388  
; FILING DATE: 9 DEC 1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: MULLINS, J.G.  
; REGISTRATION NUMBER: 33,073  
; REFERENCE/DOCKET NUMBER: 325800-332  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 201-994-1700  
; TELEFAX: 201-994-1744  
; INFORMATION FOR SEQ ID NO: 12:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 90 AMINO ACIDS  
; TYPE: AMINO ACID  
; STRANDEDNESS:  
; TOPOLOGY: LINEAR  
; MOLECULE TYPE: PROTEIN  
; US-08-464-339A-12

Query Match 71.4%; Score 30; DB 1; Length 90;  
Best Local Similarity 75.0%; Pred. No. 44;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GIGVODG 8  
DB 42 GIGVODG 49

RESULT 7  
US-08-545-809A-143  
; Sequence 143, Application US/08545809A  
; Patent No. 6096878  
; GENERAL INFORMATION:  
; APPLICANT: Honjo, Tasuku  
; APPLICANT: Matsuda, Fumihiko  
; TITLE OF INVENTION: HUMAN IMMUNOGLOBULIN VH GENE  
; NUMBER OF SEQUENCES: 145  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson, P.C.  
; STREET: 225 Franklin Street  
; CITY: Boston  
; STATE: MA  
; COUNTRY: US  
; ZIP: 02110-2804

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: Windows95  
SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/545,809A  
FILING DATE: 27-MAR-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/JP93/00603  
FILING DATE: 10-MAY-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Freeman, John W.  
REGISTRATION NUMBER: 29,066  
REFERENCE/DOCKET NUMBER: 06501/004001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-542-5070  
TELEFAX: 617-542-8906  
TELEX: 200154  
INFORMATION FOR SEQ ID NO: 143:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 116 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-545-809A-143

Query Match  
Best Local Similarity 71.4%; Score 30; DB 3; Length 116;  
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GRLVQDGC 8  
Db 27 GEGLVQPG 34

RESULT 8  
US-08-468-847B-12  
Sequence 12, Application US/08468847B  
Patent No. 5780263  
GENERAL INFORMATION:  
APPLICANT: Haslinga, Gregg A. and Adams, Mark D.  
TITLE OF INVENTION: Human CCM-Like Growth Factor  
NUMBER OF SEQUENCES: 20  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLIAN,  
STREET: 6 BECKER FARM ROAD  
CITY: ROSELAND  
STATE: NEW JERSEY  
COUNTRY: USA  
ZIP: 07068  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 INCH DISKETTE  
COMPUTER: IBM PS/2  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: WORD PERFECT 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/468,847B  
FILING DATE: 6 June 1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: MULLINS, J.G.  
REGISTRATION NUMBER: 33,073  
REFERENCE/DOCKET NUMBER: 325800-442  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 201-994-1700  
TELEFAX: 201-994-1744  
INFORMATION FOR SEQ ID NO: 12:

SEQUENCE CHARACTERISTICS:  
LENGTH: 374 AMINO ACIDS  
TYPE: AMINO ACID  
STRANDEDNESS: LINEAR  
TOPOLOGY: LINEAR  
MOLECULE TYPE: PROTEIN  
US-08-468-847B-12

Query Match  
Best Local Similarity 71.4%; Score 30; DB 1; Length 374;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GRLVQDGC 8  
Db 42 GVLVQDGC 49

RESULT 9  
US-08-459-101A-2  
Sequence 2, Application US/08459101A  
Patent No. 5945300  
GENERAL INFORMATION:  
APPLICANT: LI, ET AL.  
TITLE OF INVENTION: Connective Tissue Growth Factor-2  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLIAN,  
STREET: 6 BECKER FARM ROAD  
CITY: ROSELAND  
STATE: NEW JERSEY  
COUNTRY: USA  
ZIP: 07068  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 INCH DISKETTE  
COMPUTER: IBM PS/2  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: WORD PERFECT 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/459,101A  
FILING DATE: June 2, 1995  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US94/07736  
FILING DATE: 12 JUL 94  
ATTORNEY/AGENT INFORMATION:  
NAME: FERRARO, GREGORY D.  
REGISTRATION NUMBER: 36,134  
REFERENCE/DOCKET NUMBER: 325800-317  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 201-994-1700  
TELEFAX: 201-994-1744  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 375 AMINO ACIDS  
TYPE: AMINO ACID  
STRANDEDNESS: LINEAR  
TOPOLOGY: LINEAR  
MOLECULE TYPE: PROTEIN  
US-08-459-101A-2

Query Match  
Best Local Similarity 71.4%; Score 30; DB 2; Length 375;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GRLVQDGC 8  
Db 42 GVLVQDGC 49

RESULT 10

US-08-468-847B-11  
Sequence 11, Application US/08468847B  
Patent No. 5780263  
GENERAL INFORMATION:  
APPLICANT: Hastings, Gregg A. and Adams, Mark D.  
TITLE OF INVENTION: Human CCM-Like Growth Factor  
NUMBER OF SEQUENCES: 20  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,  
ADDRESSEE: CECCHI, STEWART & OLSTEIN  
STREET: 6 BECKER FARM ROAD  
CITY: ROSELAND  
STATE: NEW JERSEY  
COUNTRY: USA  
ZIP: 07068  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 INCH DISKETTE  
COMPUTER: IBM PS/2  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: WORD PERFECT 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/468,847B  
FILING DATE: 6 June 1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: MULLINS, J.G.  
REGISTRATION NUMBER: 33,073  
REFERENCE/DOCKET NUMBER: 325800-442  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 201-994-1700  
TELEFAX: 201-994-1744  
INFORMATION FOR SEQ ID NO: 11:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 379 AMINO ACIDS  
TYPE: AMINO ACID  
STRANDEDNESS: LINEAR  
TOPOLOGY: LINEAR  
MOLECULE TYPE: PROTEIN  
US-08-468-847B-11

Query Match 71.4%; Score 30; DB 1; Length 379;  
Best Local Similarity 75.0%; Pred. No. 2e+02;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 GGLVODG 8  
DB 42 GGLVODG 49

RESULT 11  
US-08-820-170A-34  
Sequence 34, Application US/08820170A  
Patent No. 5831058  
GENERAL INFORMATION:  
APPLICANT: Tsutomu, FUJIMURA  
APPLICANT: Takeshi, MATANABE  
APPLICANT: Masato, HORIE  
APPLICANT: Toyomasa, KATAGIRI  
TITLE OF INVENTION: HUMAN GENE  
NUMBER OF SEQUENCES: 42  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sughrue, Mion, Zinn, Macpeak & Seas  
STREET: 2100 Pennsylvania Avenue, N.W.  
CITY: Washington  
STATE: D.C.  
COUNTRY: United States  
ZIP: 20037-3202  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/820,170A  
FILING DATE:  
CLASSIFICATION: 536  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 293-7060  
TELEFAX: (202) 293-7860  
TELEX: 6491103  
INFORMATION FOR SEQ ID NO: 34:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 810 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-820-170A-34

Query Match 71.4%; Score 30; DB 2; Length 810;  
Best Local Similarity 57.1%; Pred. No. 4.4e+02;  
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 2 RGLVODG 8  
DB 201 KGLVODG 207

RESULT 12  
US-09-055-699-34  
Sequence 34, Application US/09055699  
Patent No. 6005088  
GENERAL INFORMATION:  
APPLICANT: Tsutomu, FUJIMURA  
APPLICANT: Takeshi, MATANABE  
APPLICANT: Masato, HORIE  
APPLICANT: Toyomasa, KATAGIRI  
TITLE OF INVENTION: HUMAN GENE  
NUMBER OF SEQUENCES: 42  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sughrue, Mion, Zinn, Macpeak & Seas  
STREET: 2100 Pennsylvania Avenue, N.W.  
CITY: Washington  
STATE: D.C.  
COUNTRY: United States  
ZIP: 20037-3202  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/055,699  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/820,170  
FILING DATE:  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 293-7060  
TELEFAX: (202) 293-7860  
TELEX: 6491103  
INFORMATION FOR SEQ ID NO: 34:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 810 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-055-699-34

Query Match 71.4%; Score 30; DB 3; Length 810;

Best Local Similarity 57.1%; Pred. No. 4.4e+02;  
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 2 RGLVODG 8  
:|:|:|  
Db 201 KGIODG 207

RESULT 13  
US-09-273-565-34

; Sequence 34, Application US/09273565A  
; Patent No. 6166190  
; GENERAL INFORMATION:  
; APPLICANT: FUJIMURA, TSUTOMU  
; APPLICANT: MATINABE, TAKESHI  
; APPLICANT: HORIE, MASATO  
; TITLE OF INVENTION: AN ISOLATED NUCLEIC ACID MOLECULE ENCODING HUMAN  
; FILE REFERENCE: 0-53599  
; CURRENT APPLICATION NUMBER: US/09/273,565A  
; EARLIER FILING DATE: 1999-03-22  
; EARLIER APPLICATION NUMBER: 09/055,699  
; EARLIER FILING DATE: 1998-04-07  
; EARLIER APPLICATION NUMBER: 08/820,170  
; EARLIER FILING DATE: 1997-03-19  
; EARLIER APPLICATION NUMBER: JP 63410/1996  
; EARLIER FILING DATE: 1996-03-19  
; EARLIER APPLICATION NUMBER: JP 69163/1997  
; NUMBER OF SEQ ID NOS: 95  
; SOFTWARE: Patent In Ver. 2.1  
; SEQ ID NO 34  
; LENGTH: 810  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-273-565-34

Query Match 71.4%; Score 30; DB 4; Length 810;  
Best Local Similarity 57.1%; Pred. No. 4.4e+02;  
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 2 RGLVODG 8  
:|:|:|  
Db 201 KGIODG 207

RESULT 14  
US-08-652-816A-12

; Sequence 12, Application US/08652816A  
; Patent No. 5872215  
; GENERAL INFORMATION:  
; APPLICANT: Osbourn, JK  
; APPLICANT: Allen, DJ  
; APPLICANT: McCafferty, JG  
; TITLE OF INVENTION: Specific binding members, materials and  
; NUMBER OF SEQUENCES: 33  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
; STREET: 6300 Sears Tower, 233 South Wacker Drive  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: United States of America  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/652,816A  
; FILING DATE: 23-MAY-1996  
; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: GB 9125579.4  
; FILING DATE: 02-DEC-1991

; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: GB 9125579.8

; FILING DATE: 02-DEC-1991

; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: GB 9206318.9

; FILING DATE: 24-MAR-1992

; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: GB 9206372.6

; FILING DATE: 23-SEP-1992

; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: GB 9525004.9

; FILING DATE: 07-DEC-1995

; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: GB 9610824.6

; FILING DATE: 23-MAY-1996

; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/GB92/02240

; FILING DATE: 02-DEC-1992

; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/244,597

; FILING DATE: 01-JUN-1994

; ATTORNEY/AGENT INFORMATION:  
; NAME: David W. Clough

; REGISTRATION NUMBER: 36,107

; REFERENCE/DOCKET NUMBER: 28111/33308

; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312-474-6300

; INFORMATION FOR SEQ ID NO: 12:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 118 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

US-08-652-816A-12

Query Match 69.0%; Score 29; DB 2; Length 118;  
Best Local Similarity 75.0%; Pred. No. 90;  
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 GRLVODG 8  
|||  
Db 8 GRALVOPG 15

RESULT 15  
US-08-318-157B-2

; Sequence 2, Application US/08318157B  
; Patent No. 5874540  
; GENERAL INFORMATION:  
; APPLICANT: HANSEN, Hans J.  
; APPLICANT: ARMOUR, Kathryn L.  
; TITLE OF INVENTION: CDR-GRAFTED TYPE III ANTI-CEA HUMANIZED  
; NUMBER OF SEQUENCES: 58  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 3000 K Street, N.W., Suite 500  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20007-5109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/318,157B  
; FILING DATE: 05-OCT-1994  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:

NAME: SAKE, Bernhard D.  
REGISTRATION NUMBER: 28,665  
REFERENCE/DOCKET NUMBER: 18733/464  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)672-5300  
TELEFAX: (202)672-5399  
TELEX: 904136  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 119 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-318-157B-2

Query Match 69.0%; Score 29; DB 2; Length 119;  
Best Local Similarity 75.0%; Pred. No. 90;  
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 GRLVQDG 8  
1 11111  
Db 8 CGGLVQSG 15

Search completed: June 13, 2001, 14:16:35  
Job time: 496 sec

*This Page Blank (uspto)*



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 13, 2001, 14:14:36 ; Search time 229.28 Seconds  
(without alignments)  
1.995 Million cell updates/sec

Title: PCT-US01-05825A-13

Perfect score: 47

Sequence: 1 GGCVCQPG 8

Scoring table:

BLOSUM62

Searched: Gapop 10.0 , Gapext 0.5

Total number of hits satisfying chosen parameters: 390729

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

Listing first 45 summaries

A\_Geneseq\_0401.\*  
1: /SID56/gcgdata/geneseq/geneseqp/AA1980.DAT:\*  
2: /SID56/gcgdata/geneseq/geneseqp/AA1981.DAT:\*  
3: /SID56/gcgdata/geneseq/geneseqp/AA1982.DAT:\*  
4: /SID56/gcgdata/geneseq/geneseqp/AA1983.DAT:\*  
5: /SID56/gcgdata/geneseq/geneseqp/AA1984.DAT:\*  
6: /SID56/gcgdata/geneseq/geneseqp/AA1985.DAT:\*  
7: /SID56/gcgdata/geneseq/geneseqp/AA1987.DAT:\*  
8: /SID56/gcgdata/geneseq/geneseqp/AA1988.DAT:\*  
9: /SID56/gcgdata/geneseq/geneseqp/AA1988.DAT:\*  
10: /SID56/gcgdata/geneseq/geneseqp/AA1988.DAT:\*  
11: /SID56/gcgdata/geneseq/geneseqp/AA1990.DAT:\*  
12: /SID56/gcgdata/geneseq/geneseqp/AA1991.DAT:\*  
13: /SID56/gcgdata/geneseq/geneseqp/AA1992.DAT:\*  
14: /SID56/gcgdata/geneseq/geneseqp/AA1993.DAT:\*  
15: /SID56/gcgdata/geneseq/geneseqp/AA1994.DAT:\*  
16: /SID56/gcgdata/geneseq/geneseqp/AA1995.DAT:\*  
17: /SID56/gcgdata/geneseq/geneseqp/AA1996.DAT:\*  
18: /SID56/gcgdata/geneseq/geneseqp/AA1998.DAT:\*  
19: /SID56/gcgdata/geneseq/geneseqp/AA1998.DAT:\*  
20: /SID56/gcgdata/geneseq/geneseqp/AA2000.DAT:\*  
21: /SID56/gcgdata/geneseq/geneseqp/AA2001.DAT:\*  
22: /SID56/gcgdata/geneseq/geneseqp/AA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	47	100.0	8	21	Y79117
2	44	93.6	8	21	Y79121
3	41	87.2	31	20	W88384
4	41	87.2	73	20	W88389
5	41	87.2	115	21	B41718
6	41	87.2	153	21	B41638
7	41	87.2	158	20	W88388
8	41	87.2	169	20	W88390
9	41	87.2	224	20	Y59870
10	41	87.2	247	21	Y52139
11	41	87.2	254	20	W88382

12	41	87.2	265	21	B42204	Human ORFX ORF1968
13	41	87.2	273	20	Y41759	Human PRO213-1 pro
14	41	87.2	273	20	Y41770	Human PRO1330 prot
15	41	87.2	273	20	Y41771	Human PRO1449 prot
16	41	87.2	273	21	W88381	Human neuro-growth
17	41	87.2	273	21	B44325	Human PRO213-1 pro
18	41	87.2	273	21	B44326	Human PRO1330 prot
19	41	87.2	273	21	B44327	Human PRO1449 prot
20	41	87.2	273	21	B18673	Amino acid sequenc
21	41	87.2	273	21	B18674	Amino acid sequenc
22	41	87.2	273	21	B18675	Amino acid sequenc
23	41	87.2	273	21	B24042	Human PRO1330 prote
24	41	87.2	273	21	B24043	Human PRO1449 prot
25	41	87.2	273	21	B24044	Human PRO1449 prot
26	41	87.2	273	21	B01376	Neuron-associated
27	41	87.2	273	21	Y52137	Human TANGO 125 (T
28	41	87.2	295	20	Y41685	Human PRO213 (UNQ1
29	41	87.2	295	21	B44241	Human ORFX ORF1408
30	41	87.2	307	21	B41644	Peptide antagonist
31	40	85.1	8	21	Y79125	Peptide antagonist
32	39	83.0	8	21	Y79105	Peptide antagonist
33	39	83.0	8	21	Y79118	Peptide antagonist
34	38	80.9	379	20	Y08065	Human EGF-like hom
35	38	80.9	379	20	Y13345	Amino acid sequenc
36	38	80.9	379	21	B24397	Human PRO217 prote
37	38	80.9	379	21	Y70669	Human PRO217 prote
38	38	80.9	379	21	Y44822	Human molecule ass
39	38	80.9	379	22	B53076	Human angiotensin
40	38	80.9	380	20	W88726	Secreted protein e
41	38	80.9	531	15	R48994	Human glycosyltran
42	38	80.9	531	17	R97615	Rat N-acetylglucos
43	38	80.9	531	18	W24015	Human N-acetylgluc
44	38	80.9	536	17	R97614	Rat N-acetylglucos
45	38	80.9	536	18	W24014	Rat N-acetylglucos

#### ALIGNMENTS

RESULT	1
ID	Y79117 standard; Peptide; 8 AA.
XX	
AC	Y79117:
XX	
DT	05-JUN-2000 (first entry)
XX	
DE	Peptide antagonist of zonulin.
XX	
KW	zonulin; antagonist; zonula occludens toxin receptor;
KW	blood-brain barrier; antiinflammatory; cerebroprotective;
KW	neuroprotective; dermatological; antitumor; antiviral;
KW	antibacterial; cytoskeletal; anti-HIV; vulnereary; antiallergic;
KW	hypotensive; immunosuppressive; antiparasitic; vasotropic;
KW	gastrointestinal inflammation; therapy.
XX	
OS	Synthetic.
XX	
PN	WO200007609-A1.
XX	
PD	17-FEB-2000.
XX	
PF	28-JUL-1999; 99MO-US16683.
XX	
PR	03-AUG-1998; 98US-0127815.
XX	
PA	(UYMA-) UNIV MARYLAND BALTIMORE.
XX	
PI	Fasano A;
XX	
DR	WPI, 2000-205565/18.
XX	
PT	New peptide antagonist of zonulin useful as antiinflammatory agent for

PT treating cerebral ischemia, stroke, cerebral edema, gastritis,  
 PR shigellosis, viral gastroenteritis, meningitis, encephalomyelitis  
 XX  
 PS Claim 1: Page 45; 69pp; English.

CC This present sequence is that of a peptide antagonist of zonulin  
 CC (Z), one of 25 such peptides (see Y79105-29) of the invention,  
 CC which bind to a zonula occludens toxin (ZOT) receptor, yet do not  
 CC physiologically modulate the opening of mammalian tight junctions  
 CC (TJ). The peptide antagonists are based on a common motif of ZOT  
 CC and human zonulins, which is believed to be critical for receptor  
 CC binding. They can be prepared by chemical synthesis or by use of  
 CC recombinant DNA techniques. The peptide antagonists are used as an  
 CC antiinflammatory agents in the treatment of gastrointestinal  
 CC inflammation, where they bind to the ZOT receptor in the intestine  
 CC and yet does not physiologically modulate the opening of TJ in the  
 CC increased intestinal permeability and the peptide is useful for  
 CC treating intestinal conditions that cause protein losing enteropathy  
 CC caused by infection, e.g. Clostridium difficile infection,  
 CC enterocolitis, shigellosis, viral gastroenteritis, parasite  
 CC infestation, bacterial overgrowth, whipple's disease, diseases with  
 CC mucosal erosion or ulcerations, e.g. gastritis, gastric cancer,  
 CC collagenous colitis, inflammatory bowel disease, diseases marked by  
 CC sarcoidosis lymphoma, mesenteric tuberculosis, and after surgical  
 CC correction of congenital heart disease with Fontan's operation,  
 CC mucosal diseases without ulceration, e.g. Menetrier's disease,  
 CC coeliac disease, eosinophilic gastroenteritis, and immune diseases,  
 CC e.g. systemic lupus erythematosus or food allergies, primarily to  
 CC milk.  
 CC  
 CC Sequence 8 AA;  
 CC  
 CC

Query Match 100.0%; Score 47; DB 21; Length 8;  
 Best Local Similarity 100.0%; Pred. No. 3.2e+05;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GGCVCVQPG 8  
 |||||  
 DB 1 ggcvcvqpg 8

RESULT 2

Y79121  
 Y79121 standard; Peptide; 8 AA.

XX Y79121;

DT 05-JUN-2000 (first entry)

DE Peptide antagonist of zonulin.

XX Zonulin; antagonist; zonula occludens toxin receptor;  
 KW blood-brain barrier; antiinflammatory; cerebroprotective;  
 KW neuroprotective; dermatological; anticancer; antiviral;  
 KW antibacterial; cytostatic; anti-HIV; vulnerrary; antiallergic;  
 KW hypotensive; immunosuppressive; antiparasitic; vasotrophic;  
 KW gastrointestinal inflammation; therapy.

XX Synthetic.

OS WO200007609-A1.

PN 17-FEB-2000.

PD 28-JUL-1999; 99WO-US16683.

PF 03-AUG-1998; 98US-0127815.

PR (UYMA-) UNIV MARYLAND BALTIMORE.

XX

PI Fasano A;  
 XX WPI: 2000-205565/18.  
 DR  
 XX

PT New peptide antagonist of zonulin useful as antiinflammatory agent for  
 PR treating cerebral ischemia, stroke, cerebral edema, gastritis,  
 PR shigellosis, viral gastroenteritis, meningitis, encephalomyelitis  
 XX  
 PS Claim 1: Page 46; 69pp; English.

CC This present sequence is that of a peptide antagonist of zonulin  
 CC (Z), one of 25 such peptides (see Y79105-29) of the invention,  
 CC which bind to a zonula occludens toxin (ZOT) receptor, yet do not  
 CC physiologically modulate the opening of mammalian tight junctions  
 CC (TJ). The peptide antagonists are based on a common motif of ZOT  
 CC and human zonulins, which is believed to be critical for receptor  
 CC binding. They can be prepared by chemical synthesis or by use of  
 CC recombinant DNA techniques. The peptide antagonists are used as an  
 CC antiinflammatory agents in the treatment of gastrointestinal  
 CC inflammation, where they bind to the ZOT receptor in the intestine  
 CC and yet does not physiologically modulate the opening of TJ in the  
 CC increased intestinal permeability and the peptide is useful for  
 CC treating intestinal conditions that cause protein losing enteropathy  
 CC caused by infection, e.g. Clostridium difficile infection,  
 CC enterocolitis, shigellosis, viral gastroenteritis, parasite  
 CC infestation, bacterial overgrowth, whipple's disease, diseases with  
 CC mucosal erosion or ulcerations, e.g. gastritis, gastric cancer,  
 CC collagenous colitis, inflammatory bowel disease, diseases marked by  
 CC sarcoidosis lymphoma, mesenteric tuberculosis, and after surgical  
 CC correction of congenital heart disease with Fontan's operation,  
 CC mucosal diseases without ulceration, e.g. Menetrier's disease,  
 CC coeliac disease, eosinophilic gastroenteritis, and immune diseases,  
 CC e.g. systemic lupus erythematosus or food allergies, primarily to  
 CC milk.  
 CC  
 CC Sequence 8 AA;  
 CC  
 CC

Query Match 93.6%; Score 44; DB 21; Length 8;  
 Best Local Similarity 87.5%; Pred. No. 3.2e+05;

Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GGCVCVQPG 8  
 ||:||||  
 DB 1 ggcvcvqpg 8

RESULT 3

W88384  
 W88384 standard; Peptide; 31 AA.

XX W88384;

DT 26-APR-1999 (first entry)

DE Human Zneul EGF-like domain 1.

XX Zneul-1; neuro-growth factor-like protein; human; breast cancer;  
 KW glioblastoma; pituitary adenoma; Alzheimer's disease; therapy;  
 KW nerve regeneration; hematopoiesis; fertility; contraception;  
 KW antibody; epidermal growth factor; EGF.

XX Homo sapiens.

OS WO9857983-A2.

PN 23-DEC-1998.

PD 18-JUN-1998; 98WO-US12763.

PF 18-JUN-1997; 97US-0878322.

XX

PR 18-JUN-1997; 97US-0050143.  
XX  
PA (ZYMO ) ZYMOGENETICS INC.  
XX  
PI Blumberg H, Jelinek LJ, Lehner JM, Sheppard PO;  
PI Whitmore TE;  
XX  
DR WPI; 1999-095324/08.  
XX  
PS Claim 6; Page 53; 70pp; English.  
XX  
CC This peptide comprises the first epidermal growth factor-like  
CC domain (EGF-1), i.e. amino acid residues 105-135, of human Zneul  
CC (see W88381), a new neuro-growth factor-like protein. Zneul can be  
CC used as a growth, maintenance, or differentiation factor in the  
CC spinal cord, heart, spleen, testis, thyroid and lymph nodes. It  
CC may also play a role in breast cancer, glioblastomas, and pituitary  
CC adenomas. Zneul may also be used to treat Alzheimer's disease,  
CC cancer, to repopulate blood cells after chemotherapy, to stimulate  
CC myofibroblast proliferation, stimulate or inhibit growth factors  
CC made in the placenta, in fertility and contraception, or to  
CC regenerate nerves. Claimed Zneul polypeptides (see also W88382-97),  
CC including specific domains of Zneul and epitope-bearing portions of  
CC Zneul, can be used to raise specific antibodies for use e.g. in  
CC diagnostic assays.  
XX  
SQ Sequence 31 AA;  
XX  
Query Match 87.2%; Score 41; DB 20; Length 31;  
Best Local Similarity 87.5%; Pred. No. 4.2;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 GGCVCVOPG 8  
DB 10 99gcvcvpg 17  
XX  
ID W88389 standard; Protein; 73 AA.  
XX  
AC W88389;  
XX  
DT 26-APR-1999 (first entry)  
XX  
DE Human Zneul partial polypeptide.  
XX  
KW Zneul-1; neuro-growth factor-like protein; human; breast cancer;  
KW glioblastoma; pituitary adenoma; Alzheimer's disease; therapy;  
KW nerve regeneration; haematopoiesis; fertility; contraception;  
KW antibody.  
XX  
OS Homo sapiens.  
XX  
PN MO9857983-A2.  
XX  
PD 23-DEC-1998.  
XX  
PE 18-JUN-1998; 98MO-US12763.  
XX  
PR 18-JUN-1997; 97US-087832.  
PR 18-JUN-1997; 97US-0050143.  
XX  
PA (ZYMO ) ZYMOGENETICS INC.  
XX  
PI Blumberg H, Jelinek LJ, Lehner JM, Sheppard PO;  
PI Whitmore TE;  
XX  
DR WPI; 1999-095324/08.

XX  
PT New mammalian Zneul polypeptides - used to, e.g. treat Alzheimer's  
PT disease, cancer and to repopulate blood cells  
XX  
XX Claim 6; Page 57; 70pp; English.  
XX  
CC This claimed polypeptide comprises amino acid residues 105-177 of  
CC the human Zneul mature protein (see W88382), i.e. the 2 epidermal  
CC growth factor-like domains (see also W88384-85) of Zneul. Zneul is  
CC a new neuro-growth factor-like protein that can be used as a  
CC growth, maintenance, or differentiation factor in the spinal cord,  
CC heart, spleen, testis, thyroid and lymph nodes. Zneul may also  
CC play a role in breast cancer, glioblastomas, and pituitary  
CC adenomas. Zneul can be used to treat Alzheimer's disease, cancer,  
CC to repopulate blood cells after chemotherapy, to stimulate  
CC myofibroblast proliferation, stimulate or inhibit growth factors  
CC made in the placenta, in fertility and contraception, or to  
CC regenerate nerves. Claimed Zneul polypeptides (see also W88382-97),  
CC including specific domains of Zneul and epitope-bearing portions of  
CC Zneul, can be used to raise specific antibodies for use e.g. in  
CC diagnostic assays.  
XX  
SQ Sequence 73 AA;  
XX  
Query Match 87.2%; Score 41; DB 20; Length 73;  
Best Local Similarity 87.5%; Pred. No. 8.8;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 GGCVCVOPG 8  
DB 10 99gcvcvpg 17  
XX  
ID B41718 standard; Protein; 115 AA.  
XX  
AC B41718;  
XX  
DT 08-FEB-2001 (first entry)  
XX  
DE Human ORFX ORF1482 polypeptide sequence SEQ ID NO:2964.  
XX  
KW Human; Open reading frame; ORFX; detection; cytostatic; hepatotropic;  
KW vulnery; antiparkinsonian; noctropic; neuroprotective;  
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;  
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;  
KW hypotensive; dermatological; immunosuppressive; antiinflammatory;  
KW antiviral; antibacterial; antifungal; antineumatic; antithyroid;  
KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;  
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;  
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;  
KW cholesterol ester storage; systemic lupus erythematosus; infection;  
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;  
KW allergy; aplastic anemia; nocturnal haemoglobinuria; burn; wound;  
KW bone damage; cartilage damage; antiinflammatory disease; coagulation;  
KW thrombosis; contraceptive.  
XX  
OS Homo sapiens.  
XX  
PN WO200058473-A2.  
XX  
PD 05-OCT-2000.  
XX  
PE 31-MAR-2000; 2000MO-US08621.  
XX  
PR 31-MAR-1999; 99US-0127607.  
PR 02-APR-1999; 99US-0127636.  
PR 05-APR-1999; 99US-0127728.  
PR 30-MAR-2000; 2000US-0540763.  
XX  
PA (CURA-) CURAGEN CORP.

XX Shinkets RA, Leach M;  
 XX MPI; 2000-602362/57.  
 DR N-PSDB; C75927.  
 XX  
 PT Novel nucleic acids and peptides derived from open reading frame X,  
 PT useful for treating e.g. cancers, proliferative disorders,  
 PT neurodegenerative disorders and cardiovascular disease -  
 PS Claim 11; Page 2200; 5507pp; English.  
 XX  
 CC C74446 to C77606 encode the proteins given in B40237 to B43397, which  
 CC represent the human ORFX open reading frames 1 to 3161. The ORFX  
 CC sequences have activities such as: cytostatic; hepatotropic; vulnery;  
 CC antiproliferative; antiparkinsonian; neurotropic; neuroprotective; osteopathic;  
 CC anticonvulsant; antidiabetic; immunosuppressant; immunostimulant;  
 CC cardiant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive;  
 CC dermatological; immunosuppressive; antineoplastic; antitumor;  
 CC antiviral; antifungal; antirheumatic; antidiabetic; antihypertensive;  
 CC sequences can be used for determining the presence of or predisposition  
 CC to or preventing or treating pathological conditions associated with an  
 CC ORFX-associated disorder. The nucleic acids can be used to express ORFX  
 CC proteins in gene therapy vectors. The proteins and nucleic acids may be  
 CC used to treat cancers, proliferative disorders, neurodegenerative  
 CC disorders, osteoarthritis, graft vs host disease, cardiovascular disease,  
 CC diabetes mellitus, hypertension, hypothyroidism, cholesterol ester  
 CC storage, systemic lupus erythematosus, severe combined immunodeficiency  
 CC (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune  
 CC disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and  
 CC cartilage damage, nocturnal haemoglobinuria, antineoplastic disease; to  
 CC enhance coagulation; to inhibit thrombosis; and as a contraceptive.  
 XX Sequence 115 AA;  
 SO  
 Query Match 87.2%; Score 41; DB 21; Length 115;  
 Best Local Similarity 87.5%; Pred. No. 13;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 OY 1 GGCVCVPG 8  
 II IIIII  
 Db 19 ggcvcvpg 26  
 RESULT 6  
 B41638  
 ID B41638 standard; Protein: 153 AA.  
 AC B41638;  
 XX  
 DT 08-FEB-2001 (first entry)  
 XX  
 DE Human ORFX ORF1402 polypeptide sequence SEQ ID NO:2804.  
 KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;  
 KW vulnery; antiproliferative; antiparkinsonian; neurotropic; neuroprotective;  
 KW anticonvulsant; osteopathic; antidiabetic; immunosuppressant; cardiant;  
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;  
 KW antitumor; dermatological; immunosuppressive; antineoplastic; antitumor;  
 KW antiviral; antifungal; antirheumatic; antidiabetic; antihypertensive;  
 KW antineoplastic; gene therapy; cancer; proliferative disorder; hypertension;  
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;  
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;  
 KW cholesterol ester storage; systemic lupus erythematosus; infection;  
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;  
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;  
 KW thrombosis; contraceptive; antineoplastic disease; coagulation;  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200058473-A2.

XX 05-OCT-2000.  
 PD  
 XX 31-MAR-2000; 2000MO-US08621.  
 PE  
 XX 31-MAR-1999; 9905-0127607.  
 PR 02-APR-1999; 9905-0127636.  
 PR 05-APR-1999; 9905-0127728.  
 PR 30-MAR-2000; 2000US-0540763.  
 XX  
 PA (CURA-) CURAGEN CORP.  
 XX  
 XX Shinkets RA, Leach M;  
 XX MPI; 2000-602362/57.  
 DR N-PSDB; C75847.  
 XX  
 PT Novel nucleic acids and peptides derived from open reading frame X,  
 PT useful for treating e.g. cancers, proliferative disorders,  
 PT neurodegenerative disorders and cardiovascular disease -  
 PS Claim 11; Page 2044; 5507pp; English.  
 XX  
 CC C74446 to C77606 encode the proteins given in B40237 to B43397, which  
 CC represent the human ORFX open reading frames 1 to 3161. The ORFX  
 CC sequences have activities such as: cytostatic; hepatotropic; vulnery;  
 CC antiproliferative; antiparkinsonian; neurotropic; neuroprotective; osteopathic;  
 CC anticonvulsant; antidiabetic; immunosuppressant; immunostimulant;  
 CC cardiant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive;  
 CC dermatological; immunosuppressive; antineoplastic; antitumor;  
 CC antiviral; antifungal; antirheumatic; antidiabetic; antihypertensive;  
 CC sequences can be used for determining the presence of or predisposition  
 CC to or preventing or treating pathological conditions associated with an  
 CC ORFX-associated disorder. The nucleic acids can be used to express ORFX  
 CC proteins in gene therapy vectors. The proteins and nucleic acids may be  
 CC used to treat cancers, proliferative disorders, neurodegenerative  
 CC disorders, osteoarthritis, graft vs host disease, cardiovascular disease,  
 CC diabetes mellitus, hypertension, hypothyroidism, cholesterol ester  
 CC storage, systemic lupus erythematosus, severe combined immunodeficiency  
 CC (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune  
 CC disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and  
 CC cartilage damage, nocturnal haemoglobinuria, antineoplastic disease; to  
 CC enhance coagulation; to inhibit thrombosis; and as a contraceptive.  
 XX Sequence 153 AA;  
 SO  
 Query Match 87.2%; Score 41; DB 21; Length 153;  
 Best Local Similarity 87.5%; Pred. No. 17;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 OY 1 GGCVCVPG 8  
 II IIIII  
 Db 91 ggcvcvpg 98  
 RESULT 7  
 W88388  
 ID W88388 standard; Protein: 158 AA.  
 AC W88388;  
 XX  
 DT 26-APR-1999 (first entry)  
 XX  
 DE Human Zneul partial polypeptide.  
 KW Zneul-1; neuro-growth factor-like protein; human; breast cancer;  
 KW glioblastoma; pituitary adenoma; Alzheimer's disease; therapy;  
 KW nerve regeneration; haematopoiesis; fertility; contraception;  
 KW antibody.  
 XX  
 OS Homo sapiens.  
 XX

PN W09857983-A2.  
XX  
PD 23-DEC-1998.  
XX  
PF 18-JUN-1998; 98WO-US12763.  
XX  
PR 18-JUN-1997; 97US-0878322.  
PR 18-JUN-1997; 97US-0050143.  
XX  
PA (ZYMO ) ZYMOGENETICS INC.  
XX  
PI Blumberg H, Jelinek LJ, Lehner JM, Sheppard PO;  
PI Whitmore TE;  
XX  
DR WPI; 1999-095324/08.  
XX  
PT New mammalian Zneul polypeptides - used to, e.g. treat Alzheimer's  
PT disease, cancer and to repopulate blood cells  
XX  
PS Claim 6; Page 56; 70pp; English.  
XX  
XX This claimed polypeptide comprises amino acid residues 1-158 of the  
CC human Zneul mature protein (see W88382), i.e. the first HSMC3W5A  
CC homology domain and first EGF-like domain of Zneul. Zneul is a new  
CC neuro-growth factor-like protein that can be used as a growth,  
CC maintenance, or differentiation factor in the spinal cord, heart,  
CC spleen, testis, thyroid and lymph nodes. Zneul may also play a  
CC role in breast cancer, glioblastomas, and pituitary adenomas.  
CC Zneul can be used to treat Alzheimer's disease, cancer, to  
CC repopulate blood cells after chemotherapy, to stimulate  
CC myofibroblast proliferation, stimulate or inhibit growth factors  
CC made in the placenta, in fertility and contraception, or to  
CC regenerate nerves. Claimed Zneul polypeptides (see also W88382-97),  
CC including specific domains of Zneul and epitope-bearing portions of  
CC Zneul, can be used to raise specific antibodies for use e.g. in  
CC diagnostic assays.  
XX  
SQ Sequence 158 AA;

Query Match 87.2%; Score 41; DB 20; Length 158;  
Best Local Similarity 87.5%; Pred. No. 17;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGVCVQPG 8  
DB 95 ggcvcvpg 102

RESULT 8  
W88390  
ID W88390 standard; Protein; 169 AA.  
XX  
AC W88390;  
XX  
DT 26-APR-1999 (first entry)  
XX  
DE Human Zneul partial polypeptide.  
XX  
XX Zneul-1; neuro-growth factor-like protein; human; breast cancer;  
KW glioblastoma; pituitary adenoma; Alzheimer's disease; therapy;  
KW nerve regeneration; haematopoiesis; fertility; contraception;  
KM antibody.  
XX  
OS Homo sapiens.  
XX  
PN W09857983-A2.  
XX  
PD 23-DEC-1998.  
XX  
PF 18-JUN-1998; 98WO-US12763.  
XX  
PR 18-JUN-1997; 97US-0878322.  
XX

PR 18-JUN-1997; 97US-0050143.  
XX  
XX (ZYMO ) ZYMOGENETICS INC.  
PA  
PI Blumberg H, Jelinek LJ, Lehner JM, Sheppard PO;  
PI Whitmore TE;  
XX  
DR WPI; 1999-095324/08.  
XX  
PT New mammalian Zneul polypeptides - used to, e.g. treat Alzheimer's  
PT disease, cancer and to repopulate blood cells  
XX  
PS Claim 6; Page 57-58; 70pp; English.  
XX  
XX This claimed polypeptide comprises amino acid residues 105-273 of  
CC the human Zneul mature protein (see W88382), i.e. the 2 epidermal  
CC growth factor-like domains (see also W88384-85) and the second  
CC HSMC3W5A homology domain of Zneul. Zneul is a new neuro-growth  
CC factor-like protein that can be used as a growth, maintenance, or  
CC differentiation factor in the spinal cord, heart, spleen, testis,  
CC thyroid and lymph nodes. Zneul may also play a role in breast  
CC cancer, glioblastomas, and pituitary adenomas. Zneul can be used  
CC to treat Alzheimer's disease, cancer, to repopulate blood cells  
CC after chemotherapy, to stimulate myofibroblast proliferation,  
CC stimulate or inhibit growth factors made in the placenta, in  
CC fertility and contraception, or to regenerate nerves. Claimed  
CC Zneul polypeptides (see also W88382-97), including specific domains  
CC of Zneul and epitope-bearing portions of Zneul, can be used to  
CC raise specific antibodies for use e.g. in diagnostic assays.  
XX  
SQ Sequence 169 AA;

Query Match 87.2%; Score 41; DB 20; Length 169;  
Best Local Similarity 87.5%; Pred. No. 18;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGVCVQPG 8  
DB 10 ggcvcvpg 17

RESULT 9  
Y59870  
ID Y59870 standard; Protein; 224 AA.  
XX  
AC Y59870;  
XX  
DT 19-JAN-2000 (first entry)  
XX  
DE Human normal uterus tissue derived protein 33.  
XX  
XX Human; uterus; cancer; treatment; anticancer; cytostatic; gene therapy;  
KW EGF; expressed sequence tag.  
XX  
OS Homo sapiens.  
XX  
PN DE19817946-A1.  
XX  
PD 21-OCT-1999.  
XX  
PF 17-APR-1998; 98DE-1017946.  
XX  
PR 17-APR-1998; 98DE-1017946.  
XX  
PA (META-) METAGEN GES GENOMFORSCHUNG MBH.  
XX  
PI Rosenthal A, Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E;  
PI WPI; 1999-591956/51.  
XX  
DR N-PSDB; 241336.  
XX  
PT New nucleic acid sequences expressed in normal uterine tissues, and

Y52138-Y52140 are amino acid sequences of spliced variants of TANGO 125 (T125). The T125 protein (Y52137) has two epidermal growth factor

New mammalian zneui polypeptides - used to, e.g. treat Alzheimer's disease

XX PS Claim 6; Page 48-49; 70pp; English.

XX CC This polypeptide comprises human zneul mature polypeptide. Zneul

CC is a new neuro-growth factor-like protein (see also W88381). Its

CC closest human homologue is HSMHC3W5A, a gene in the HLA class III

CC region, which is contained in a cosmid which contains Notch 4.

CC Zneul is also homologous to Notch 4 in its EGF-like domains and may

CC be involved in EGF receptor pathways. Zneul is widely expressed in

CC adult tissues, with high expression in heart, placenta, spleen,

CC testis, thymoid, spinal cord and lymph node. Zneul can be used as

CC a growth, maintenance, or differentiation factor in the spinal

CC cord, heart, spleen, testis, thymoid and lymph nodes. It may

CC play a role in breast cancer, glioblastomas, and pituitary

CC adenomas. Zneul may also be used to treat Alzheimer's disease,

CC cancer, to repopulate blood cells after chemotherapy, to stimulate

CC myofibroblast proliferation, stimulate or inhibit growth factors

CC made in the placenta, in fertility and contraception, or to

CC regenerate nerves. Claimed zneul polypeptides (see also W88382-97),

CC including specific domains of zneul and epitope-bearing portions of

CC zneul, can be used to raise specific antibodies for use e.g. in

CC diagnostic assays.

XX SQ Sequence 254 AA;

Query Match 87.28; Score 41; DB 20; Length 254;

Best Local Similarity 87.5%; Pred. No. 26;

Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GGCVCVOPG 8

Db 95 ggcvcvpg 102

II IIIII

RESULT 12

ID B42204 standard; Protein: 265 AA.

XX AC B42204;

XX DT 08-FEB-2001 (first entry)

XX DE Human ORFX ORF1968 polypeptide sequence SEQ ID NO:3936.

XX KW Human: open reading frame; ORFX: detection; cytostatic; hepatotropic;

KW vulnery; antiproliferative; antiparkinsonian; neuroprotective;

KW anticonvulsant; osteopathic; antichratic; immunosuppressant; cardiac;

KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;

KW hypotensive; dermatological; immunosuppressive; antiinflammatory;

KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;

KW antineoplastic; gene therapy; cancer; proliferative disorder; hypertension;

KW neurodegenerative disorder; osteoarthritis; graft vs host disease;

KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;

KW cholesterol ester storage; systemic lupus erythematosus; infection;

KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;

KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;

KW bone damage; cartilage damage; antiinflammatory disease; coagulation;

KW thrombosis; contraceptive.

XX OS Homo sapiens.

XX PN WO200058473-A2.

XX PD 05-OCT-2000.

XX PF 31-MAR-2000; 2000MO-US08621.

XX PR 31-MAR-1999; 99US-0127607.

PR 02-APR-1999; 99US-0127636.

PR 05-APR-1999; 99US-0127728.

PR 30-MAR-2000; 2000US-0540763.

PA (CURA-) CURAGEN CORP.

XX PI Shimkets RA, Leach M;

XX DR WPI: 2000-602362/57.

XX DR N-PSDB; C76413.

XX PT Novel nucleic acids and peptides derived from open reading frame X,

PT useful for treating e.g. cancers, proliferative disorders,

PT neurodegenerative disorders and cardiovascular disease -

XX PS Claim 11; Page 3102; 5507pp; English.

XX CC C74446 to C77606 encode the proteins given in B40237 to B43397, which

CC represent the human ORFX open reading frames 1 to 3161. The ORFX

CC sequences have activities such as: cytostatic; hepatotropic; vulnery;

CC antiproliferative; antiparkinsonian; neurotropic; neuroprotective; osteopathic;

CC anticonvulsant; antichratic; immunosuppressant; immunostimulant;

CC cardiant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive;

CC dermatological; immunosuppressive; antirheumatic; antithyroid; antineoplastic;

CC antiviral; antifungal; antirheumatic; antithyroid; and antineoplastic. The

CC sequences can be used for determining the presence of or predisposition

CC to, or preventing or treating pathological conditions associated with an

CC ORFX-associated disorder. The nucleic acids can be used to express ORFX

CC proteins in gene therapy vectors. The proteins and nucleic acids may be

CC used to treat cancers, proliferative disorders, neurodegenerative

CC disorders, osteoarthritis, graft vs host disease, cardiovascular disease,

CC diabetes mellitus, hypertension, hypothyroidism, cholesterol ester

CC storage, systemic lupus erythematosus, severe combined immunodeficiency

CC (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune

CC disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and

CC cartilage damage, nocturnal haemoglobinuria, antiinflammatory disease; to

CC enhance coagulation; to inhibit thrombosis; and as a contraceptive.

XX SQ Sequence 265 AA;

Query Match 87.28; Score 41; DB 21; Length 265;

Best Local Similarity 87.5%; Pred. No. 27;

Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GGCVCVOPG 8

Db 114 ggcvcvpg 121

II IIIII

RESULT 13

ID Y41769 standard; Protein: 273 AA.

XX AC Y41769;

XX DT 07-DEC-1999 (first entry)

XX DE Human PRO213-1 protein sequence.

XX KW Human: PRO; EST: expressed sequence tag; PCR primer; hybridisation;

KW probe; blood coagulation disorder; cancer; cellular adhesion disorder;

KW secreted protein; transmembrane protein.

XX OS Homo sapiens.

XX PN WO946281-A2.

XX PD 16-SEP-1999.

XX PF 08-MAR-1999; 99MO-US05028.

XX PR 10-MAR-1998; 98US-0077450.

PR 11-MAR-1998; 98US-0077632.

PR 11-MAR-1998; 98US-0077641.

PR 11-MAR-1998; 98US-0077649.

PR 12-MAR-1998; 98US-0077791.





PR	12-MAR-1996	98US-0077791
PR	13-MAR-1996	98US-0078000
PR	17-MAR-1996	98US-0040224
PR	20-MAR-1996	98US-0078816
PR	20-MAR-1996	98US-0078860
PR	20-MAR-1996	98US-0078939
PR	20-MAR-1996	98US-0078939
PR	25-MAR-1996	98US-0079294
PR	26-MAR-1996	98US-0079656
PR	27-MAR-1996	98US-0079663
PR	27-MAR-1996	98US-0079664
PR	27-MAR-1996	98US-0079668
PR	27-MAR-1996	98US-0079728
PR	27-MAR-1996	98US-0079786
PR	30-MAR-1996	98US-0079920
PR	30-MAR-1996	98US-0079923
PR	31-MAR-1996	98US-0080105
PR	31-MAR-1996	98US-0080107
PR	31-MAR-1996	98US-0080194
PR	01-APR-1996	98US-0080327
PR	01-APR-1996	98US-0080328
PR	01-APR-1996	98US-0080334
PR	01-APR-1996	98US-0080334
PR	08-APR-1996	98US-0081049
PR	08-APR-1996	98US-0081070
PR	08-APR-1996	98US-0081071
PR	09-APR-1996	98US-0081195
PR	09-APR-1996	98US-0081203
PR	09-APR-1996	98US-0081229
PR	15-APR-1996	98US-0081617
PR	15-APR-1996	98US-0081638
PR	15-APR-1996	98US-0081952
PR	15-APR-1996	98US-0081955
PR	21-APR-1996	98US-0082568
PR	21-APR-1996	98US-0082569
PR	22-APR-1996	98US-0082700
PR	22-APR-1996	98US-0082704
PR	22-APR-1996	98US-0082804
PR	23-APR-1996	98US-0082767
PR	23-APR-1996	98US-0082796
PR	27-APR-1996	98US-0083322
PR	28-APR-1996	98US-0083326
PR	29-APR-1996	98US-0083392
PR	29-APR-1996	98US-0083495
PR	29-APR-1996	98US-0083496
PR	29-APR-1996	98US-0083559
PR	30-APR-1996	98US-0083742
PR	05-MAY-1996	98US-0084366
PR	06-MAY-1996	98US-0084414
PR	06-MAY-1996	98US-0084441
PR	07-MAY-1996	98US-0084598
PR	07-MAY-1996	98US-0084600
PR	07-MAY-1996	98US-0084627
PR	07-MAY-1996	98US-0084639
PR	07-MAY-1996	98US-0084640
PR	13-MAY-1996	98US-0085328
PR	13-MAY-1996	98US-0085338
PR	15-MAY-1996	98US-0085339
PR	15-MAY-1996	98US-0085573
PR	15-MAY-1996	98US-0085579
PR	15-MAY-1996	98US-0085580
PR	15-MAY-1996	98US-0085582
PR	15-MAY-1996	98US-0085689
PR	15-MAY-1996	98US-0085697
PR	15-MAY-1996	98US-0085700

PR	15-MAY-1998;	98US-0085704.
PR	18-MAY-1998;	98US-0086023.
PR	22-MAY-1998;	98US-0086392.
PR	22-MAY-1998;	98US-0086414.
PR	22-MAY-1998;	98US-0086430.
PR	22-MAY-1998;	98US-0086486.
PR	28-MAY-1998;	98US-0087098.
PR	28-MAY-1998;	98US-0087106.
PR	28-MAY-1998;	98US-0087208.
PR	30-JUL-1998;	98US-0094651.
PR	11-SEP-1998;	98US-0100038.
XX	(GETH ) GENENTECH INC.	
XX		
PI	Wood WI, Goddard A, Gurney A, Yuan J, Baker KP, Chen J;	
XX		
DR	WPI, 1999-551358/46.	
DR	N-PSDB; Z34312.	
PT	New secreted and transmembrane polypeptides and their polynucleotides,	
XX	useful for treating blood coagulation disorders, cancers and cellular	
XX	adhesion disorders -	
PS	Claim 12; Fig 215; 530pp; English.	
XX		
CC	The present invention describes secreted and transmembrane polypeptides	
CC	and their polynucleotides. The nucleotide sequences are useful as	
CC	sources of probes, primers, for chromosome mapping, and for generation	
CC	of antisense sequences. They can also be used to create transgenic	
CC	animals. The proteins can be used to treat a variety of diseases and	
CC	disorders, depending on their function. Diseases that may be treated	
CC	include blood coagulation disorders, cancers and cellular adhesion	
CC	disorders. They may also be used to raise antibodies. Z33891 to	
CC	Z34338, and Y41685 to Y41774 represent polynucleotide and polypeptide	
CC	sequence given in the exemplification of the present invention.	
XX		
SQ	Sequence    273 AA:	
	Query Match                 87.2%; Score 41; DB 20; Length 273;	
	Best Local Similarity      87.5%; Pred. No. 28;	
	Matches    7; Conservative          1; Indels        0; Gaps        0;	
OY	1 GGCVCPG 8	
Db	114 ggcvcvpg 121	
RESULT 15		
Y41771		
ID	Y41771 standard; Protein: 273 AA.	
XX		
AC	Y41771;	
DT	07-DEC-1999 (first entry)	
XX		
DE	Human PRO1449 protein sequence.	
XX		
KW	Human; PRO; EST; expressed sequence tag; PCR primer; hybridisation;	
KW	probe; blood coagulation disorder; cancer; cellular adhesion disorder;	
XX	secreted protein; transmembrane protein.	
OS	Homo sapiens.	
XX		
PN	WO9946281-A2.	
XX		
PD	16-SEP-1999.	
XX		
PF	08-MAR-1999; 99WO-US05028.	
XX		
XX	10-MAR-1998; 98US-0077450.	
PR	11-MAR-1998; 98US-0077632.	
PR	11-MAR-1998; 98US-0077641.	

```

PR 11-MAR-1998; 98US-0077649.
PR 12-MAR-1998; 98US-0077791.
PR 13-MAR-1998; 98US-0078004.
PR 17-MAR-1998; 98US-0040220.
PR 20-MAR-1998; 98US-0078886.
PR 20-MAR-1998; 98US-0078910.
PR 20-MAR-1998; 98US-0078936.
PR 20-MAR-1998; 98US-0078939.
PR 25-MAR-1998; 98US-0079294.
PR 26-MAR-1998; 98US-0079656.
PR 27-MAR-1998; 98US-0079663.
PR 27-MAR-1998; 98US-0079664.
PR 27-MAR-1998; 98US-0079689.
PR 27-MAR-1998; 98US-0079728.
PR 30-MAR-1998; 98US-0079786.
PR 30-MAR-1998; 98US-0079920.
PR 31-MAR-1998; 98US-0080105.
PR 31-MAR-1998; 98US-0080107.
PR 31-MAR-1998; 98US-0080165.
PR 31-MAR-1998; 98US-0080194.
PR 01-APR-1998; 98US-0080327.
PR 01-APR-1998; 98US-0080328.
PR 01-APR-1998; 98US-0080333.
PR 01-APR-1998; 98US-0080334.
PR 08-APR-1998; 98US-0081049.
PR 08-APR-1998; 98US-0081070.
PR 08-APR-1998; 98US-0081071.
PR 09-APR-1998; 98US-0081195.
PR 09-APR-1998; 98US-0081203.
PR 15-APR-1998; 98US-0081229.
PR 15-APR-1998; 98US-0081817.
PR 15-APR-1998; 98US-0081838.
PR 15-APR-1998; 98US-0081952.
PR 15-APR-1998; 98US-0081955.
PR 21-APR-1998; 98US-0082568.
PR 21-APR-1998; 98US-0082569.
PR 22-APR-1998; 98US-0082700.
PR 22-APR-1998; 98US-0082704.
PR 22-APR-1998; 98US-0082804.
PR 23-APR-1998; 98US-0082767.
PR 23-APR-1998; 98US-0082796.
PR 27-APR-1998; 98US-0083336.
PR 28-APR-1998; 98US-0083322.
PR 28-APR-1998; 98US-0083392.
PR 28-APR-1998; 98US-0083495.
PR 29-APR-1998; 98US-0083495.
PR 29-APR-1998; 98US-0083499.
PR 29-APR-1998; 98US-0083500.
PR 29-APR-1998; 98US-0083545.
PR 29-APR-1998; 98US-0083554.
PR 29-APR-1998; 98US-0083558.
PR 30-APR-1998; 98US-0083559.
PR 05-MAY-1998; 98US-0083742.
PR 06-MAY-1998; 98US-0084366.
PR 06-MAY-1998; 98US-0084414.
PR 06-MAY-1998; 98US-0084441.
PR 07-MAY-1998; 98US-0084598.
PR 07-MAY-1998; 98US-0084600.
PR 07-MAY-1998; 98US-0084627.
PR 07-MAY-1998; 98US-0084637.
PR 07-MAY-1998; 98US-0084639.
PR 07-MAY-1998; 98US-0084640.
PR 13-MAY-1998; 98US-0084643.
PR 13-MAY-1998; 98US-0085323.
PR 13-MAY-1998; 98US-0085338.
PR 13-MAY-1998; 98US-0085339.
PR 15-MAY-1998; 98US-0085573.
PR 15-MAY-1998; 98US-0085579.
PR 15-MAY-1998; 98US-0085580.
PR 15-MAY-1998; 98US-0085582.
PR 15-MAY-1998; 98US-0085689.
PR 15-MAY-1998; 98US-0085697.

```

```

PR 15-MAY-1998; 98US-0085700.
PR 15-MAY-1998; 98US-0085704.
PR 18-MAY-1998; 98US-0086023.
PR 22-MAY-1998; 98US-0086392.
PR 22-MAY-1998; 98US-0086410.
PR 22-MAY-1998; 98US-0086430.
PR 22-MAY-1998; 98US-0086486.
PR 28-MAY-1998; 98US-0087098.
PR 28-MAY-1998; 98US-0087106.
PR 28-MAY-1998; 98US-0087208.
PR 30-JUL-1998; 98US-0094651.
PR 11-SEP-1998; 98US-0100038.

```

(GETH ) GENENTECH INC.

Wood WI, Goddard A, Gurney A, Yuan J, Baker KP, Chen J;

WPI; 1999-551358/46.

N-PSDB; 234313.

New secreted and transmembrane polypeptides and their polynucleotides, useful for treating blood coagulation disorders, cancers and cellular adhesion disorders -

Claim 12; Fig 217; 530pp; English.

CC The present invention describes secreted and transmembrane polypeptides and their polynucleotides. The nucleotide sequences are useful as sources of probes, primers, for chromosome mapping, and for generation of antisense sequences. They can also be used to create transgenic animals. The proteins can be used to treat a variety of diseases and disorders, depending on their function. Diseases that may be treated include blood coagulation disorders, cancers and cellular adhesion disorders. They may also be used to raise antibodies. 233891 to 234338, and 141685 to 141774 represent polynucleotide and polypeptide sequence given in the exemplification of the present invention.

XX Sequence 273 AA;

Query Match 87.28; Score 41; DB 20; Length 273;

Best Local Similarity 87.3%; Pred. No. 28;

Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGVCVQPG 8

Db 114 ggcvcvpg 121

Search completed: June 13, 2001, 14:14:36  
Job time: 377 sec



*This Page Blank (uspto)*

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

## OM protein - protein search, using sw model

Run on: June 13, 2001, 14:10:44 ; Search time 130.61 Seconds  
(Without alignments)  
4.209 Million cell updates/sec

Title: PCT-US01-05825A-13

Perfect score: 47

Sequence: 1 GGVQVQPG 8

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 198801 seqs, 68722935 residues

Total number of hits satisfying chosen parameters: 198801

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR-67:\*  
2: p1r1:\*  
3: p1r2:\*  
4: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	38	80.9	278	2 E65020	ethanolamine utili
2	38	80.9	300	1 B70627	hypothetical prote
3	38	80.9	379	2 A59180	Wnt inhibitory fac
4	38	80.9	531	2 JN0586	beta-1,4-mannosyl-
5	38	80.9	536	2 A43415	beta-1,4-mannosyl-
6	38	80.9	536	2 JIC462	beta-1,4-mannosyl-
7	37	78.7	98	2 S26929	Ig heavy chain V r
8	37	78.7	120	2 S44111	Ig heavy chain V-D
9	36	76.6	378	2 B59180	glutamate dehydrog
10	35	74.5	421	1 A38168	hypothetical prote
11	35	74.5	467	2 D72461	peptide synthase -
12	35	74.5	532	2 T34917	Integrin beta-3 ch
13	35	74.5	680	2 P00510	Integrin beta-3 ch
14	35	74.5	723	2 P00509	Integrin beta-3 ch
15	35	74.5	780	2 T03156	ribonucleoside-dip
16	35	74.5	799	2 S55656	ribonucleoside-dip
17	34	72.3	92	2 D37057	epithelial cell gl
18	34	72.3	145	2 H69051	heterodisulfide re
19	34	72.3	196	2 F71525	hypothetical prote
20	34	72.3	196	2 H81681	mat protein TC0628
21	34	72.3	250	2 T10512	hypothetical prote
22	34	72.3	264	2 S12027	methionyl aminopep
23	34	72.3	267	2 S44228	strf protein - Str
24	34	72.3	281	2 S17776	hypothetical prote
25	34	72.3	423	2 S39830	hypothetical prote
26	34	72.3	427	2 D83347	probable aminotran
27	34	72.3	444	2 T15310	hypothetical prote
28	34	72.3	468	2 D70532	hypothetical prote
29	34	72.3	589	2 T08466	carbonate dehydrat

30	34	72.3	753	2 B36268	platelet glycoprot
31	34	72.3	788	2 A60798	platelet glycoprot
32	34	72.3	788	2 A26547	platelet glycoprot
33	34	72.3	788	2 I77349	platelet glycoprot
34	34	72.3	1827	2 T34288	hypothetical prote
35	34	72.3	2380	2 T29551	hypothetical prote
36	34	71.3	881	2 T31818	hypothetical prote
37	33	70.2	96	2 D83228	hypothetical prote
38	33	70.2	116	1 A1H02P	Ig heavy chain V-I
39	33	70.2	256	2 T16805	hypothetical prote
40	33	70.2	266	2 S71025	lipopolysaccharide
41	33	70.2	403	2 E69873	cell-division prot
42	33	70.2	495	2 A39900	alanine transamina
43	33	70.2	495	2 A40465	alanine transamina
44	33	70.2	505	2 T37975	probable alanine a
45	33	70.2	505	2 H83196	glycerol kinase PA

## ALIGNMENTS

RESULT 1  
E65020  
ethanolamine utilization protein EutJ - Escherichia coli (strain K-12)  
C:Species: Escherichia coli  
C:Date: 12-Sep-1997 #sequence\_revision 17-Sep-1997 #text\_change 08-Oct-1999  
R:Accession: E65020  
R:Blatner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;  
.A.; Rose, D.J.; Mau, B.; Shao, Y.  
Science 277, 1453-1462, 1997  
A>Title: The complete genome sequence of Escherichia coli K-12.  
A:Reference number: A64720; MUID:97426617  
A:Accession: E65020  
A:Status: Preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-278 <BLAT>  
A:Cross-references: GB:A6000332; GB:U00096; NID:g1788789; PIDN:AAC75507.1; PID:g17887  
A:Experimental source: strain K-12, substrain MG1655  
A:Genetics: eutJ  
A:Gene: eutJ

Query Match  
Best Local Similarity 80.9%; Score 38; DB 2; Length 278;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 GGVQVQPG 8  
DB 229 GGSCHQPG 236

RESULT 2  
B70627  
hypothetical protein RV0263c - Mycobacterium tuberculosis (strain H37Rv)  
C:Species: Mycobacterium tuberculosis  
C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 21-Jul-2000  
R:Accession: B70627  
R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon  
R.; Connor, R.; Davies, R.; Devlin, R.; Felwell, T.; Gentles, S.; Hamlin, N.; Holroyd,  
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skellton, S.; Squares, S.  
Nature 393, 537-544, 1998  
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A>Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno  
A:Reference number: A70500; MUID:98295987  
A:Accession: B70627  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-300 <COB>  
A:Cross-references: GB:Z86089; GB:AL123456; NID:g3261711; PIDN:CAB06690.1; PID:g18501  
A:Experimental source: strain H37Rv  
A:Genetics:  
A:Gene: RV0263c  
C:Superfamily: hypothetical protein H11730

Query Match  
Best Local Similarity 80.9%; Score 38; DB 1; Length 300;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GGVGVOP 7  
DB 120 GGVGVOP 126

RESULT 3  
A:Accession: A59180  
Mnt inhibitory factor-1 - human  
C:Species: Homo sapiens (man)  
C:Date: 18-Feb-2000 #sequence\_revision 18-Feb-2000 #text\_change 18-Feb-2000  
C:Accession: A59180  
R:Hisish, J.C.; Kodjabachian, L.; Rebbert, M.L.; Ratner, A.; Smallwood, P.M.; Samos, C.H.  
Nature 398, 431-436, 1999  
A:Title: A new secreted protein that binds to Wnt proteins and inhibits their activities  
A:Reference number: A59180; MUID:99215557  
A:Accession: A59180  
A:Status: preliminary; not compared with conceptual translation  
A:Molecule type: mRNA  
A:Cross-references: 1-379 <HSI>  
A:Cross-references: GB:AF122922; NID:94585369; PIDN:AAD25402.1; PID:94585370

Query Match  
Best Local Similarity 80.9%; Score 38; DB 2; Length 379;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GGVGVOP 8  
DB 221 GGLCVTPG 228

RESULT 4  
A:Accession: JN0586  
beta-1,4-mannosyl-glycoprotein beta-1,4-N-acetylglucosaminyltransferase (EC 2.4.1.144)  
N:Alternate names: beta-D-mannoside beta1,4-N-acetylglucosaminyltransferase III; N-acetylglucosaminyltransferase (EC 2.4.1.144)  
C:Species: Homo sapiens (man)  
C:Date: 24-Feb-1994 #sequence\_revision 24-Feb-1994 #text\_change 20-Jun-2000  
C:Accession: JN0586  
R:Iihara, Y.; Nishikawa, A.; Tohma, T.; Soejima, H.; Nishikawa, N.; Taniguchi, N.  
J. Biochem. 113, 692-698, 1993  
A:Title: CDNA cloning, expression, and chromosomal localization of human N-acetylglucosaminyltransferase number: JN0586; MUID:93380894  
A:Reference number: JN0586; MUID:93380894  
A:Accession: JN0586  
A:Molecule type: mRNA  
A:Residues: 1-531 <TMA>  
A:Cross-references: GB:D13789; NID:9398137; PIDN:BA02937.1; PID:9398138  
A:Experimental source: fetal liver  
C:Comment: This enzyme catalyzes the addition of N-acetylglucosamine in beta1-4 linkage  
C:Genetics:  
A:Map position: 22q13.1  
C:Superfamily: UDP-N-acetylglucosamine acyltransferase  
C:Keywords: glycoprotein; glycosyltransferase; hexosyltransferase; transmembrane protein  
F:6-21/Domain: transmembrane #status predicted <TMA>  
F:159-180/Region: epidermal growth factor-like  
F:139,239,257,395/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match  
Best Local Similarity 80.9%; Score 38; DB 2; Length 531;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 GGVGVOP 8  
DB 112 GGVGVOP 119

RESULT 5  
A:Accession: A43415  
beta-1,4-mannosyl-glycoprotein beta-1,4-N-acetylglucosaminyltransferase (EC 2.4.1.144)  
C:Species: Rattus norvegicus (Norway rat)  
C:Date: 10-Jun-1993 #sequence\_revision 18-Nov-1994 #text\_change 20-Jun-2000  
C:Accession: A43415  
R:Nishikawa, A.; Ihara, Y.; Hatakeyama, M.; Kangawa, K.; Taniguchi, N.  
J. Biol. Chem. 267, 18199-18204, 1992  
A:Title: Purification, CDNA cloning, and expression of UDP-N-acetylglucosamine: beta-1,4-mannosyltransferase  
A:Reference number: A43415; MUID:92388193  
A:Accession: A43415  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-536 <NIS>  
A:Cross-references: GB:D10852; NID:9220821; PIDN:BA01625.1; PID:9220822  
A:Note: sequence extracted from NCBI backbone (NCBI:112852, NCBI:112855)  
C:Superfamily: UDP-N-acetylglucosamine acyltransferase  
C:Keywords: glycoprotein; glycosyltransferase; hexosyltransferase; transmembrane prot

Query Match  
Best Local Similarity 80.9%; Score 38; DB 2; Length 536;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 GGVGVOP 8  
DB 112 GGVGVOP 119

RESULT 6  
A:Accession: JN0586  
beta-1,4-mannosyl-glycoprotein beta-1,4-N-acetylglucosaminyltransferase (EC 2.4.1.144)  
N:Alternate names: GlcNAc-TIII; glycosyltransferase  
C:Species: Mus musculus (house mouse)  
C:Date: 08-Jan-1996 #sequence\_revision 08-Feb-1996 #text\_change 18-Jun-1999  
C:Accession: JN0586  
R:Bhaumik, M.; Seidlin, M.F.; Stanley, P.  
Gene 164, 295-300, 1995  
A:Title: Cloning and chromosomal mapping of the mouse Mgat3 gene encoding N-acetylglucosaminyltransferase number: JN0586; MUID:96069598  
A:Reference number: JN0586; MUID:96069598  
A:Accession: JN0586  
A:Molecule type: mRNA  
A:Residues: 1-536 <BHA>  
A:Cross-references: GB:J39373; NID:9100872; PIDN:AAB71422.1; PID:9100873  
A:Experimental source: liver, 129 SV  
C:Comment: The action of this enzyme correlates with tissue specific functions, and is synthesized by mammalian cells.  
C:Genetics:  
A:Gene: Mgat3  
A:Map position: 15  
C:Superfamily: UDP-N-acetylglucosamine acyltransferase  
C:Keywords: glycoprotein; glycosyltransferase; hexosyltransferase; liver; transmembrane  
F:6-21/Domain: transmembrane #status predicted <TMA>  
F:243,261,399/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match  
Best Local Similarity 80.9%; Score 38; DB 2; Length 536;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 GGVGVOP 8  
DB 112 GGVGVOP 119

RESULT 7  
A:Accession: S26929  
Ig heavy chain V region (DP-33) - human (fragment)  
C:Species: Homo sapiens (man)  
C:Date: 22-Nov-1993 #sequence\_revision 10-Nov-1995 #text\_change 23-Jul-1999  
C:Accession: S26929  
R:Tomlinson, I.M.; Walter, G.; Marks, J.D.; Llewellyn, M.B.; Winter, G.  
J. Mol. Biol. 227, 776-798, 1992

A:Title: The repertoire of human germline V(H) sequences reveals about fifty groups of V  
 A:Reference number: S26885; MUID:93021117  
 A:Accession: S26929  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-98 <TOM>  
 A:Cross-references: EMBL:212335; NID:932889; PIDN:CAA78205.1; PID:932890  
 C:Superfamily: Immunoglobulin V region; Immunoglobulin homology  
 C:Keywords: heterotrimer; Immunoglobulin  
 F:15-98/Domain: Immunoglobulin homology <IMM>

Query Match 78.7%; Score 37; DB 2; Length 98;  
 Best Local Similarity 87.5%; Pred. No. 5.8;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GGCVCVQPG 8  
 ||| ||||  
 DB 8 GGVVQVPG 15

RESULT 8  
 S4411  
 Ig heavy chain V-D-J region - human  
 C:Species: Homo sapiens (man)  
 C:Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 23-Jul-1999  
 C:Accession: S44111  
 R:Hawkins, R.E.; Zhu, D.; Ovecke, M.; Winter, G.; Hamblin, T.J.; Stevenson, F.K.  
 Submitted to the EMBL Data Library, March 1994  
 A:Description: Idiotypic vaccination against human B-cell lymphoma: rescue of variable r  
 A:Reference number: S44105  
 A:Accession: S44111  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-120 <HAW>  
 A:Cross-references: EMBL:231387; NID:9472965; PIDN:CAA83262.1; PID:9940522  
 C:Superfamily: Immunoglobulin V region; Immunoglobulin homology  
 C:Keywords: heterotrimer; Immunoglobulin  
 F:15-98/Domain: Immunoglobulin homology <IMM>

Query Match 78.7%; Score 37; DB 2; Length 120;  
 Best Local Similarity 87.5%; Pred. No. 7;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GGCVCVQPG 8  
 ||| ||||  
 DB 8 GGVVQVPG 15

RESULT 9  
 B59180  
 Wnt inhibitory factor-1 - zebra fish  
 C:Species: Brachydanio rerio (zebra fish)  
 C:Date: 18-Feb-2000 #sequence\_revision 18-Feb-2000 #text\_change 18-Feb-2000  
 C:Accession: B59180  
 R:Hasleh, J.C.; Kojabachian, L.; Rebett, M.L.; Ratner, A.; Smallwood, P.M.; Samos, C.H.  
 Nature 398, 431-436, 1999  
 A:Title: A new secreted protein that binds to Wnt proteins and inhibits their activities  
 A:Reference number: A59180; MUID:99215557  
 A:Accession: B59180  
 A:Status: preliminary; not compared with conceptual translation  
 A:Molecule type: mRNA  
 A:Residues: 1-378 <HST>  
 A:Cross-references: GB:AF122925; NID:94585375; PIDN:AA025405.1; PID:94585376  
 C:Genetics:  
 A:Gene: WIF-1

Query Match 76.6%; Score 36; DB 2; Length 378;  
 Best Local Similarity 62.5%; Pred. No. 30;  
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 GGCVCVQPG 8  
 ||:|: ||  
 DB 219 GGCKMSPG 226

RESULT 10  
 A38168  
 glutamate dehydrogenase (EC 1.4.1.2) - Peptostreptococcus asaccharolyticus  
 N:Alternate names: glutamic dehydrogenase; NAD-specific glutamate dehydrogenase  
 C:Species: Peptostreptococcus asaccharolyticus  
 C:Date: 16-Oct-1992 #sequence\_revision 23-Mar-1995 #text\_change 11-Jun-1999  
 C:Accession: A38168  
 R:Snedecor, B.; Chu, H.; Chen, E.  
 J. Bacteriol. 173, 6162-6167, 1991  
 A:Title: Selection, expression, and nucleotide sequencing of the glutamate dehydrogen  
 A:Reference number: A38168; MUID:92011378  
 A:Accession: A38168  
 A:Molecule type: DNA  
 A:Residues: 1-421 <SNE>  
 A:Cross-references: GB:M76403; NID:9150669; PIDN:AAA25611.1; PID:9150670  
 C:Superfamily: glutamate dehydrogenase (NAD(P)+)  
 C:Keywords: NAD; oxidoreductase  
 F:106/Binding site: substrate (Lys) #status predicted

Query Match 74.5%; Score 35; DB 1; Length 421;  
 Best Local Similarity 71.4%; Pred. No. 50;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 GGCVCVQPG 7  
 ||:|: |  
 DB 107 GGICVDP 113

RESULT 11  
 D72461  
 hypothetical protein APE2336 - Aeropyrum pernix (strain K1)  
 C:Species: Aeropyrum pernix  
 C:Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 20-Aug-1999  
 C:Accession: D72461  
 R:Kawarayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Ta  
 awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.  
 DNA Res. 6, 83-101, 1999  
 A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aero  
 A:Reference number: A72450; MUID:99310339  
 A:Accession: D72461  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-467 <KAW>  
 A:Cross-references: DBJ:AP000064; NID:95105945; PIDN:BAAB1348.1; PID:01045134; PID:9  
 A:Experimental source: strain K1  
 C:Genetics:  
 A:Gene: APE2336

Query Match 74.5%; Score 35; DB 2; Length 467;  
 Best Local Similarity 75.0%; Pred. No. 55;  
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 GGCVCVQPG 8  
 || | |||  
 DB 115 GGECRQPG 122

RESULT 12  
 T34917  
 peptide synthase - Streptomyces coelicolor  
 C:Species: Streptomyces coelicolor  
 C:Date: 05-Nov-1999 #sequence\_revision 05-Nov-1999 #text\_change 18-Aug-2000  
 C:Accession: T34917  
 R:Oliver, K.; Harris, D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.  
 submitted to the EMBL Data Library, January 1998  
 A:Reference number: Z21558

A:Accession: T34917  
 A:Status: Preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-532 <OLI>  
 A:Cross-references: EMBL:AL021409; PIDN:CAI6182.1; GSPDB:GN00070; SCOEDB:SC3F7.11  
 A:Experimental source: strain A3(2)  
 C:Genetics:  
 A:Gene: SCOEDB:SC3F7.11  
 C:Superfamily: 4-coumarate--CoA ligase; acetate--CoA ligase homology  
 F:51-516/Domain: acetate--CoA ligase homology <ACL>

Query Match  
 Best Local Similarity 74.5%; Score 35; DB 2; Length 532;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 GVCVOPG 8  
 |||||  
 Db 376 GVCVTPG 382

## RESULT 13

Integrin beta-3 chain - mouse (fragment)  
 C:Species: Mus musculus (house mouse)  
 C:Date: 03-Feb-1994 #sequence\_revision 03-Feb-1994 #text\_change 23-Jul-1999  
 C:Accession: PNO510  
 R:Cleatut, A.M.; Rosa, J.P.; Letourneur, F.; Poncz, M.; Rifat, S.  
 Biochem. Biophys. Res. Commun. 193, 771-778, 1993  
 A:Title: A comparative analysis of cDNA-derived sequences for rat and mouse beta-3 integ  
 A:Reference number: PNO509; MUID:93290675  
 A:Accession: PNO510  
 A:Status: nucleic acid sequence not shown  
 A:Molecule type: mRNA  
 A:Residues: 1-680 <CIE>  
 C:Superfamily: Integrin beta chain; laminin-type EGF-like homology  
 C:Keywords: cell adhesion; duplication; heterodimer; membrane protein

Query Match  
 Best Local Similarity 74.5%; Score 35; DB 2; Length 680;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 VCVOPG 8  
 |||||  
 Db 509 VCVOPG 514

## RESULT 14

Integrin beta-3 chain - rat (fragment)  
 C:Species: Rattus norvegicus (Norway rat)  
 C:Date: 03-Feb-1994 #sequence\_revision 03-Feb-1994 #text\_change 23-Jul-1999  
 C:Accession: PNO509  
 R:Cleatut, A.M.; Rosa, J.P.; Letourneur, F.; Poncz, M.; Rifat, S.  
 Biochem. Biophys. Res. Commun. 193, 771-778, 1993  
 A:Title: A comparative analysis of cDNA-derived sequences for rat and mouse beta-3 integ  
 A:Reference number: PNO509; MUID:93290675  
 A:Accession: PNO509  
 A:Status: nucleic acid sequence not shown  
 A:Molecule type: mRNA  
 A:Residues: 1-723 <CIE>  
 C:Superfamily: Integrin beta chain; laminin-type EGF-like homology  
 C:Keywords: cell adhesion; duplication; heterodimer; membrane protein

Query Match  
 Best Local Similarity 74.5%; Score 35; DB 2; Length 723;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 VCVOPG 8  
 |||||  
 Db 548 VCVOPG 553

## RESULT 15

ribonucleoside-diphosphate reductase (EC 1.17.4.1) large chain - alcelaphine herpesv  
 T03156  
 N:Alternate names: ribonucleotide reductase large chain  
 C:Species: alcelaphine herpesvirus 1  
 C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 21-Jul-2000  
 C:Accession: T03156  
 R:Ensser, A.; Pflanz, R.; Fleckenstein, B.  
 J. Virol. 71, 6517-6525, 1997  
 A:Title: Primary structure of the alcelaphine herpesvirus 1 genome.  
 A:Reference number: Z14840; MUID:97404659  
 A:Accession: T03156  
 A:Status: translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-780 <ENS>  
 A:Cross-references: EMBL:AF005370; NID:92337967; PIDN:AAC58108.1; PID:92338024  
 C:Complex: dimer; large and small chain  
 C:Superfamily: herpesvirus ribonucleoside-diphosphate reductase large chain  
 C:Keywords: deoxyribonucleotide biosynthesis; early protein; oxidoreductase

Query Match  
 Best Local Similarity 74.5%; Score 35; DB 2; Length 780;  
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 GVCVOPG 8  
 |||||  
 Db 472 GRCVTPG 479

Search completed: June 13, 2001, 14:10:45  
 Job time: 146 sec





**This Page Blank (uspto)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 13, 2001, 14:21:45 ; Search time 74.44 Seconds

(without alignments)  
3.681 Million cell updates/sec

Title: PCT-US01-05825A-13

Perfect score: 47

Sequence: 1 GGCVCQPG 8

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 3425486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : swissprot\_39:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	38	80.9	278	1	EUTJ_ECOLI
2	38	80.9	279	1	EUTJ_SALTY
3	38	80.9	379	1	WIFI_HUMAN
4	38	80.9	379	1	WIFI_MOUSE
5	38	80.9	531	1	GNT3_HUMAN
6	38	80.9	536	1	GNT3_MOUSE
7	38	80.9	536	1	GNT3_MOUSE
8	35	74.5	421	1	DHE2_MOUSE
9	35	74.5	787	1	ITB3_MOUSE
10	35	72.3	263	1	AMPW_SALTY
11	34	72.3	281	1	STRF_STRGR
12	34	72.3	423	1	SHP1_YEAST
13	34	72.3	468	1	STHA_MYCTU
14	34	72.3	589	1	CAH_DUNSA
15	34	72.3	788	1	ITB3_HUMAN
16	34	72.3	867	1	EF2_BLAHO
17	33	70.2	116	1	HY3Q_HUMAN
18	33	70.2	266	1	LIZB_HAERIN
19	33	70.2	332	1	MC4R_RAT
20	33	70.2	403	1	TEI2_HUMAN
21	33	70.2	403	1	YLAO_BACSU
22	33	70.2	495	1	ALAT_HUMAN
23	33	70.2	495	1	ALAT_HUMAN
24	33	70.2	505	1	ALAT_RAT
25	33	70.2	505	1	ALAT_SCHPO
26	33	70.2	570	1	GLPK_PSEAE
27	33	70.2	598	1	DCPY_NEUCR
28	33	70.2	600	1	DPO2_MOUSE
29	32	68.1	30	1	PG20_AGRBL
30	32	68.1	341	1	VP3_GRLV
31	32	68.1	431	1	NFIH_CHICK
32	32	68.1	441	1	NFIH_HUMAN
33	32	68.1	441	1	NFIH_MESAV

34	32	68.1	441	1	NFIH_MOUSE
35	32	68.1	509	1	NFIH_RAT
36	32	68.1	522	1	NFIH_CHICK
37	32	68.1	573	1	ERGL_RAT
38	32	68.1	687	1	GALX_KIULU
39	32	68.1	1183	1	DRPL_RAT
40	32	68.1	1185	1	DRPL_HUMAN
41	32	68.1	2318	1	NPC3_MOUSE
42	31	66.0	88	1	PER_DROTE
43	31	66.0	91	1	Y4BH_RHISN
44	31	66.0	114	1	PER_DROOR
45	31	66.0	130	1	KR3A_SHEEP

## ALIGNMENTS

RESULT 1  
ID: EUTJ\_ECOLI STANDARD: PRT: 278 AA.  
AC P77277;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE ETHANOLAMINE UTILIZATION PROTEIN EUTJ.  
GN EUTJ.  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OX Escherichia.  
NCBI\_Taxid=562;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-K12 / MG1655;  
RX MEDLINE-97426617; PubMed-9278503;  
RA Blactner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
RA Mau B., Shao Y.,  
RT "The complete genome sequence of Escherichia coli K-12";  
RL Science 277:1453-1474(1997).  
RM [2]  
RN SEQUENCE FROM N.A.  
RP STRAIN-K12;  
RC Alba H., Bada T., Fujita K., Hayashi K., Honjo A., Horuchi T.,  
RA Ikemoto K., Inada T., Isono K., Itoh T., Kanai K., Kasai H.,  
RA Kasai H., Kim S., Kimura S., Kitagawa M., Kitakawa M., Makino K.,  
RA Masuda S., Miki T., Mizobuchi K., Mori H., Motomura K., Nakamura Y.,  
RA Nishimoto H., Nishio Y., Oshima T., Saito N., Sempel G., Seki Y.,  
RA Tagami H., Takemoto K., Wada C., Yamamoto Y., Yano M.;  
RA Submitted (Jan-1997) to the EMBL/Genbank/DBJ databases.  
RL -I- PATHWAY: ETHANOLAMINE UTILIZATION.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
CC EMBL: AE000332; AAC75507.1; -  
CC DR EMBL: D90874; BAA16332.1; -  
CC DR EMBL: E01484; eutJ.  
CC ECOCENE: B01484; eutJ.  
CC DR ECOCENE: 278 AA; 30082 MW; 8F92B55DE54ED9FF CRC64;  
CC SQ  
Query Match 80.9%; Score 38; DB 1; Length 278;  
Best Local Similarity 75.0%; Pred No. 5.4;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```

RESULT 2
ID EUTJ_SALTY STANDARD; PRT: 279 AA.
AC 041794;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DE 30-MAY-2000 (Rel. 39, Last annotation update)
GN ETHANOLAMINE UTILIZATION PROTEIN EUTJ.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_Taxid=602;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 14028S;
RA MEDLINE=95173114; PubMed=7868611;
RT "Ethanolamine utilization in Salmonella typhimurium: nucleotide
sequence, protein expression, and mutational analysis of the ccha
cchB gene and its regulatory cluster."
RL J. Bacteriol. 177:1357-1366(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=LT2;
RA MEDLINE=9335039; PubMed=10464203;
RT "The 17-gene ethanolamine (eut) operon of Salmonella typhimurium
encodes five homologues of carboxysome shell proteins."
RL J. Bacteriol. 181:5317-5329(1999).
RN [3]
RP ETHANOLAMINE UTILIZATION.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U018560; AAA80210.1; -
DR EMBL; AF093749; AAC78119.1; -
DR StyGene; SG10554; eutJ.
SQ SEQUENCE 279 AA; 30018 MW; 28B8CD89141D8D90 CRC64;

Query Match
Best Local Similarity 80.9%; Score 38; DB 1; Length 279;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 GGCVCVPG 8
DB 230 GGCVCVPG 237

```

```

RA Hsieh J.-C., Kodjabachian L., Rebert M.L., Ratner A.,
RA Smallwood P.M., Samos C.H., Nusse R., David I.B., Nathans J.;
RT "A new secreted protein that binds to Wnt proteins and inhibits their
RT activities."
RL Nature 398:431-436(1999).
CC -!- FUNCTION: BINDS TO WNT PROTEINS AND INHIBITS THEIR ACTIVITIES. MAY
CC BE INVOLVED IN MESODERM SEGMENTATION.
CC -!- SUBCELLULAR LOCATION: SECRETED.
CC -!- SIMILARITY: CONTAINS 5 EGF-LIKE DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF122922; AAD25402.1; -
DR MIM; 605186; -
DR HSP; P00743; ICCF.
DR InterPro; IPR000561; -
DR InterPro; IPR002049; -
DR Pfam; PF00008; EGF_5.
DR PRINTS; PR00011; EGFAMININ.
DR PROSITE; PS00022; EGF_1; 5.
DR PROSITE; PS01186; EGF_2; 4.
KW Repeat; EGF-like domain; Signal; Developmental protein.
FT CHAIN 1 28
FT STAGN 1 28
FT STAGN 29 379
FT STAGN 177 208
FT DOMAIN 209 240
FT DOMAIN 241 272
FT DOMAIN 273 304
FT DOMAIN 305 336
FT DOMAIN 336 366
FT DISULFID 177 186
FT DISULFID 182 192
FT DISULFID 198 200
FT DISULFID 209 218
FT DISULFID 214 224
FT DISULFID 230 232
FT DISULFID 241 250
FT DISULFID 246 256
FT DISULFID 262 264
FT DISULFID 273 282
FT DISULFID 278 288
FT DISULFID 294 296
FT DISULFID 305 314
FT DISULFID 310 320
FT DISULFID 326 328
FT CARBOHYD 88
FT CARBOHYD 245
SQ SEQUENCE 379 AA; 41512 MW; 27782370A266E784 CRC64;

Query Match
Best Local Similarity 80.9%; Score 38; DB 1; Length 379;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 GGCVCVPG 8
DB 221 GGCVCVPG 228

```

```

RESULT 4
ID WIF1_MOUSE STANDARD; PRT: 379 AA.
AC 09WDA1;
DT 01-OCT-2000 (Rel. 40, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DE 01-OCT-2000 (Rel. 40, Last annotation update)
DE WNT INHIBITORY FACTOR 1 PRECURSOR (WIF-1).
GN WIF1.
MEDLINE=92315557; PubMed=10201374;

```

OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99215557; PubMed=10201374;  
RA Hsieh J.-C., Kodjabachian L., Redbert M.L., Rattner A.,  
RA Smallwood P.M., Samos C.H., Nusse R., David I.B., Nathans J.;  
RT "A new secreted protein that binds to Wnt proteins and inhibits their  
RT activities";  
RL Nature 398:431-436(1999).  
CC -I- FUNCTION: BINDS TO WNT PROTEINS AND INHIBITS THEIR ACTIVITIES. MAY  
CC BE INVOLVED IN MESODERM SEGMENTATION.  
CC -I- SUBCELLULAR LOCATION: SECRETED.  
CC -I- TISSUE SPECIFICITY: EXPRESSION HIGHEST IN HEART AND LUNG. LOWER IN  
CC BRAIN AND EYE.  
CC -I- SIMILARITY: CONTAINS 5 EGF-LIKE DOMAINS.  
CC -----  
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: AF122923; AAD25403.1; -  
DR HSP: P00740; 11XA.  
DR MGD: MG1:134433; M1F1.  
DR InterPro: IPR000561; -  
DR InterPro: IPR002049; -  
DR Pfam: PF00008; EGF\_5.  
DR PRINTS: PR00011; EGF-LAMININ.  
DR PROSITE: PS00022; EGF\_1; 5.  
DR PROSITE: PS01186; EGF\_2; 4.  
KW Repeat: EGF-like domain; Signal: Developmental protein.  
FT SIGNAL 1 28  
FT CHAIN 29 379  
FT DOMAIN 177 208  
FT DOMAIN 209 240  
FT DOMAIN 241 272  
FT DOMAIN 273 304  
FT DOMAIN 305 336  
FT DISULFID 177 186  
FT DISULFID 182 192  
FT DISULFID 198 200  
FT DISULFID 209 218  
FT DISULFID 214 224  
FT DISULFID 230 232  
FT DISULFID 241 250  
FT DISULFID 246 256  
FT DISULFID 262 264  
FT DISULFID 273 282  
FT DISULFID 278 288  
FT DISULFID 294 296  
FT DISULFID 305 314  
FT DISULFID 310 320  
FT DISULFID 326 328  
FT CARBOHYD 88  
FT CARBOHYD 245  
SQ SEQUENCE 379 AA; 41590 MW; E375F2642B2BC9A CRC64;

Query Match 80.9%; Score 38; DB 1; Length 379;  
Best Local Similarity 75.0%; Pred. No. 7.1;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 GGCVCVPG 8  
DB 221 GGLCVTPG 228

RESULT 5  
GNT3\_HUMAN  
ID GNT3\_HUMAN STANDARD; PRT; 531 AA.  
AC G09327;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 01-OCT-1996 (Rel. 34, Last annotation update)  
DE BETA-1,4-MANNOSYL-GLYCOPROTEIN BETA-1,4-N-ACETYLGLUCOSAMINYL-  
DE TRANSFERASE (EC 2.4.1.144) (N-GLYCOSYL-OLIGOSACCHARIDE-GLYCOPROTEIN  
DE N-ACETYLGLUCOSAMINYLTRANSFERASE III) (N-ACETYLGLUCOSAMINYLTRANSFERASE  
DE III) (GNT-III) (GLCNAc-T III).  
GN MGAT3 OR GGN3.  
OS Homo sapiens (human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=93380894; PubMed=8370666;  
RA Ihara Y., Nishikawa A., Toma T., Soejima H., Nishikawa N., Taniguchi N.;  
RT "cDNA cloning, expression, and chromosomal localization of human N-  
RT acetylglucosaminyltransferase III (Gnt-III).";  
RL J. Biochem. 113:692-698(1993).  
CC -I- FUNCTION: IT IS INVOLVED IN THE REGULATION OF THE BIOSYNTHESIS AND  
CC BIOLOGICAL FUNCTION OF GLYCOPROTEIN OLIGOSACCHARIDES. CATALYZES  
CC THE ADDITION OF N-ACETYLGLUCOSAMINE IN BETA 1-4 LINKAGE TO THE  
CC BETA-LINKED MANNOSE OF THE TRIMANNOSE CORE OF N-LINKED SUGAR  
CC CHAINS. IT IS ONE OF THE MOST IMPORTANT ENZYMES INVOLVED IN THE  
CC REGULATION OF THE BIOSYNTHESIS OF GLYCOPROTEIN OLIGOSACCHARIDES.  
CC -I- CATALYTIC ACTIVITY: UDP-N-ACETYL-D-GLUCOSAMINE + N-ACETYL-BETA-D-  
CC GLUCOSAMINYL-1,2-ALPHA-D-MANNOSYL-1,3-(N-ACETYL-BETA-D-  
CC GLUCOSAMINYL-1,2-ALPHA-D-MANNOSYL-1,6)-BETA-D-MANNOSYL-1,4-N-  
CC ACETYL-BETA-D-GLUCOSAMINYL-R = UDP + N-ACETYL-BETA-D-GLUCOSAMINYL-  
CC 1,2-ALPHA-D-MANNOSYL-1,3-(N-ACETYL-BETA-D-GLUCOSAMINYL-1,2-ALPHA-  
CC D-MANNOSYL-1,6)-(N-ACETYL-BETA-D-GLUCOSAMINYL-1,4)-BETA-D-  
CC MANNOSEYL-1,4-N-ACETYL-BETA-D-GLUCOSAMINYL-R.  
CC -I- PATHWAY: GLYCOSYLATION  
CC -I- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. GOLGI.  
CC -----  
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: D13789; BAA02937.1; -  
DR PIR: JN0586; JN0586.  
DR MIM: 604621; -  
KW Glycosyltransferase; Transmembrane; Signal-anchor;  
KW Glycoprotein; Golgi stack.  
FT DOMAIN 1 5  
FT TRANSMEM 6 21  
FT SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN).  
FT DOMAIN 22 531  
FT LUMENAL, CATALYTIC (POTENTIAL).  
FT DOMAIN 33 84  
FT CARBOHYD 139 139  
FT CARBOHYD 239 239  
FT CARBOHYD 257 257  
FT CARBOHYD 395 395  
SQ SEQUENCE 531 AA; 61026 MW; ED657ECC06CADC CRC64;

Query Match 80.9%; Score 38; DB 1; Length 531;  
Best Local Similarity 75.0%; Pred. No. 9.5;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 GGCVCVPG 8  
DB 112 GGCVCVPG 119

RESULT 6

GNT3\_MOUSE  
ID GNT3\_MOUSE STANDARD: PRT: 536 AA.  
AC Q10470: P70386;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE BETA-1,4-MANNOSYL-GLYCOPROTEIN BETA-1,4-N-ACETYLGLUCOSAMINYL-  
TRANSFERASE (EC 2.4.1.144) (N-GLYCOSYL-OLIGOSACCHARIDE-GLYCOPROTEIN  
DE N-ACETYLGLUCOSAMINYLTRANSFERASE III) (N-ACETYLGLUCOSAMINYLTRANSFERASE  
DE III) (GNT-III) (GLCNAc-T III).  
GN MGAT3.  
OS Mus musculus (Mouse).  
CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OK NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-129/SV; TISSUE=Liver;  
RX MEDLINE=96069598; PubMed=7590346;  
RA Bhannik M., Seidn M.F., Stanley P.;  
RT Cloning and chromosomal mapping of the mouse Mgat3 gene encoding N-  
acetylglucosaminyltransferase III.";  
RL Gene 164:295-300(1995).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-129/SV; TISSUE=Liver;  
RX MEDLINE=97214999; PubMed=9061364;  
RA Platel J.J., Sarkar M., Schachter H., March J.D.;  
RT Isolation, characterization and inactivation of the mouse Mgat3 gene:  
the bisecting N-acetylglucosamine in asparagine-linked  
RT oligosaccharides appears dispensable for viability and reproduction.";  
RL Glycobiology 7:45-56(1997).  
CC -1- FUNCTION: IT IS INVOLVED IN THE REGULATION OF THE BIOSYNTHESIS AND  
BIOLOGICAL FUNCTION OF GLYCOPROTEIN OLIGOSACCHARIDES. CATALYZES  
CC THE ADDITION OF N-ACETYLGLUCOSAMINE IN BETA 1-4 LINKAGE TO THE  
CC BETA-LINKED MANNOSE OF THE TRIMANNOSE CORE OF N-LINKED SUGAR  
CC CHAINS. IT IS ONE OF THE MOST IMPORTANT ENZYMES INVOLVED IN THE  
CC REGULATION OF THE BIOSYNTHESIS OF GLYCOPROTEIN OLIGOSACCHARIDES.  
CC -1- CATALYTIC ACTIVITY: UDP-N-ACETYL-D-GLUCOSAMINE + N-ACETYL-BETA-D-  
GLUCOSAMINYL-1,2-ALPHA-D-MANNOSYL-1,3-(N-ACETYL-BETA-D-  
ACETYL-BETA-D-GLUCOSAMINYL-R = UDP + N-ACETYL-BETA-D-GLUCOSAMINYL-  
1,2-ALPHA-D-MANNOSYL-1,3-(N-ACETYL-BETA-D-GLUCOSAMINYL-1,2-ALPHA-  
D-MANNOSYL-1,6)-(N-ACETYL-BETA-D-GLUCOSAMINYL-1,4)-BETA-D-  
MANNOsyl-1,4-N-ACETYL-BETA-D-GLUCOSAMINYL-R.  
CC -1- PATHWAY: GLYCOSYLATION.  
CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. GOLGI.  
CC -1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN BRAIN AND KIDNEY AND TO A  
CC MUCH LESSER EXTENT IN STOMACH, HEART, INTESTINE, UTERUS, TESTIS,  
CC OVARY AND LUNG. NOT PRESENT IN SPLEEN, LIVER AND MUSCLE.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
DR EMBL: L39373; AA071422.1;  
DR EMBL: U66844; AAC53064.1; ALT\_INIT.  
DR MGD: MG1:104532; Mgat3.  
KW Transferase; Glycosyltransferase; Transmembrane; Signal-anchor;  
KW Glycoprotein; Golgi stack.  
FT DOMAIN 1 5  
FT TRANSMEM 6 21  
FT DOMAIN 22 536  
FT DOMAIN 33 84  
FT CAROHD 243 243  
FT CAROHD 261 261  
FT CAROHD 399 399  
FT VARIANT 44 44  
FT VARIANT 64 64  
D -> G (IN 129/SV).  
D -> G (IN 129/SV).

FT VARIANT 135 135  
FT VARIANT 137 137  
FT VARIANT 241 241  
FT VARIANT 524 524  
FT VARIANT 534 534  
SO SEQUENCE 536 AA; 61743 MW; F9FD5283867F2994 CRC64;  
  
Query Match  
Best Local Similarity 80.9%; Score 38; DB 1; Length 536;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
  
Oy 1 GGYCOPG 8  
Db 112 GGYCOPG 119  
  
RESULT 7  
GNT3\_RAT  
ID GNT3\_RAT STANDARD: PRT: 536 AA.  
AC 002527;  
DT 01-FEB-1994 (Rel. 28, Created)  
DT 01-FEB-1994 (Rel. 28, Last sequence update)  
DT 01-OCT-1996 (Rel. 34, Last annotation update)  
DE BETA-1,4-MANNOSYL-GLYCOPROTEIN BETA-1,4-N-ACETYLGLUCOSAMINYL-  
TRANSFERASE (EC 2.4.1.144) (N-GLYCOSYL-OLIGOSACCHARIDE-GLYCOPROTEIN  
DE N-ACETYLGLUCOSAMINYLTRANSFERASE III) (N-ACETYLGLUCOSAMINYLTRANSFERASE  
DE III) (GNT-III) (GLCNAc-T III).  
GN MGAT3 OR GNT3.  
OS Rattus norvegicus (Rat).  
CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OK NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A., AND SEQUENCE OF 285-295; 445-451 AND 492-507.  
RC STRAIN-DONRYU; TISSUE=Kidney;  
RX MEDLINE=92388193; PubMed=1325461;  
RA Nishikawa A., Ihara Y., Hatakeyama M., Kangawa K., Taniguchi N.;  
RT Purification, cDNA cloning, and expression of  
RT UDP-N-acetylglucosamine: beta-D-mannoside  
RT beta-1,4-n-acetylglucosaminyltransferase III from rat kidney.";  
RL J. Biol. Chem. 267:18199-18204(1992).  
CC -1- FUNCTION: IT IS INVOLVED IN THE REGULATION OF THE BIOSYNTHESIS AND  
BIOLOGICAL FUNCTION OF GLYCOPROTEIN OLIGOSACCHARIDES. CATALYZES  
CC THE ADDITION OF N-ACETYLGLUCOSAMINE IN BETA 1-4 LINKAGE TO THE  
CC BETA-LINKED MANNOSE OF THE TRIMANNOSE CORE OF N-LINKED SUGAR  
CC CHAINS. IT IS ONE OF THE MOST IMPORTANT ENZYMES INVOLVED IN THE  
CC REGULATION OF THE BIOSYNTHESIS OF GLYCOPROTEIN OLIGOSACCHARIDES.  
CC -1- CATALYTIC ACTIVITY: UDP-N-ACETYL-D-GLUCOSAMINE + N-ACETYL-BETA-D-  
GLUCOSAMINYL-1,2-ALPHA-D-MANNOSYL-1,3-(N-ACETYL-BETA-D-  
ACETYL-BETA-D-GLUCOSAMINYL-R = UDP + N-ACETYL-BETA-D-GLUCOSAMINYL-  
1,2-ALPHA-D-MANNOSYL-1,3-(N-ACETYL-BETA-D-GLUCOSAMINYL-1,2-ALPHA-  
D-MANNOSYL-1,6)-(N-ACETYL-BETA-D-GLUCOSAMINYL-1,4)-BETA-D-  
MANNOsyl-1,4-N-ACETYL-BETA-D-GLUCOSAMINYL-R.  
CC -1- PATHWAY: GLYCOSYLATION.  
CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. GOLGI.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
DR EMBL: D10852; BA01625.1;  
DR PIR: A43415; A43415.  
DR TRANSFERASE: Glycosyltransferase; Transmembrane; Signal-anchor;  
KW Glycoprotein; Golgi stack.  
FT DOMAIN 1 5  
FT TRANSMEM 6 21  
FT DOMAIN 22 536  
LUMENAL, CATALYTIC (POTENTIAL).  
CYTOPLASMIC (POTENTIAL).  
SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN).

FT	DOMAIN	23	84	PRO-RICH.
FT	CAROHAD	243	243	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CAROHAD	261	261	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CAROHAD	399	399	N-LINKED (GLCNAC. . .) (POTENTIAL).
SO	SEQUENCE	536 AA:	61/62 MW;	81817BEB6C307438 CRC64;

  

Query Match	Best Local Similarity	80.9%;	Score 38;	DB 1;	Length 536;
Matches	6;	Conservative	1;	Mismatches	1;
				Indels	0;
				Gaps	0;

  

OY	1	GGCVQPG	8
DB	112	GGCVCKPG	119

  

RESULT	8
DHE2_PEPAS	STANDARD: PRT: 421 AA.
ID	DHE2_PEPAS
AC	P28997;
DT	01-DEC-1992 (Rel. 24, Created)
DT	01-DEC-1992 (Rel. 24, Last sequence update)
DT	01-OCT-1996 (Rel. 34, Last annotation update)
DE	NAD-SPECIFIC GLUTAMATE DEHYDROGENASE (EC 1.4.1.2) (NAD-GDH).
OC	Peptostreptococcus asaccharolyticus (Peptococcus asaccharolyticus).
OC	Bacteria; Filicutes; Bacillus/Clostridium group; Clostridiaceae;
OC	Peptostreptococcus.
OX	NCBI_TaxID=1258;
RM	[1]
RP	SEQUENCE FROM N.A.
RX	MEDLINE-92011378; PubMed-1917850;
RA	Snedecor B., Chu H., Chen E.Y.;
RT	"Selection, expression, and nucleotide sequencing of the glutamate dehydrogenase gene of Peptostreptococcus asaccharolyticus.";
RL	J. Bacteriol. 173:6162-6167(1991).
CC	-1 CATALYTIC ACTIVITY: L-GLUTAMATE + H(2)O + NAD(+) = 2-OXOGLUTARATE + NH(3) + NADH.
CC	-1 PATHWAY: FIRST STEP IN THE HYDROXYGLUTARATE PATHWAY (ROUTE FOR THE ENERGY-YIELDING GLUTAMATE FERMENTATION).
CC	-1 SUBUNIT: HOMOHEXAMER.
CC	-1 SIMILARITY: BELONGS TO THE GLU/LEU/PHE/VAL DEHYDROGENASES FAMILY.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <a href="http://www.isb-sib.ch/announce/sib">http://www.isb-sib.ch/announce/sib</a> ).
CC	or send an email to <a href="mailto:license@sib-sb.ch">license@sib-sb.ch</a> ).
CC	-----
DR	EMBL: M76403; AAA25611.1; -
DR	PIR: A38168; A38168.
DR	HSSP: P80319; 1GTM.
DR	InterPro: IPR001625; -
DR	Pfam: PF00208; GLFV_dehydrog; 1.
DR	PRINTS: PRO0082; GLFDHGRNASE.
DR	PROSITE: PS00074; GLFV_DEHYDROGENASE; 1.
DR	Oxidoreductase; NAD.
FT	ACT_SITE 106 106
SO	SEQUENCE 421 AA: 46514 MW; DDA64241C027422C CRC64;

  

Query Match	Best Local Similarity	74.5%;	Score 35;	DB 1;	Length 421;
Matches	5;	Conservative	1;	Mismatches	1;
				Indels	0;
				Gaps	0;

  

OY	1	GGCVQVP	7
DB	107	GGICVDP	113

  

RESULT	9
FTB3_MOUSE	

ID	ENTRY	STANDARD	PROT	787 AA
AD	ITB3_MOUSE			
AC	054890:			
DT	15-DEC-1998 (Rel. 37, Created)			
DT	15-DEC-1998 (Rel. 37, Last sequence update)			
DT	15-JUL-1999 (Rel. 38, Last annotation update)			
DE	PLATELET MEMBRANE GLYCOPROTEIN IIAA PRECURSOR (GPIIa) (INTEGRIN BETA-3) (CD61).			
GN	ITGB3.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX	NCBI_Taxid=10090;			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=C3H/HEN;			
RA	McHugh K.P., Teitelbaum S.L., Kitazawa S., Ross F.P.:			
RL	Submitted (SEP-1997) to the EMBL/Genbank/DBJ databases.			
CC	-1- RECEPTOR FOR ADHESIVE PROTEINS. IT BINDS FIBRINOGEN, FIBRONECTIN, VITRONECTIN AND VWF. INTEGRINS RECOGNIZE THE SEQUENCE R-G-D IN THEIR LIGAND.			
CC	-1- SUBUNIT: DIMER OF AN ALPHA AND BETA SUBUNIT. BETA-3 IS EITHER ASSOCIATED WITH ALPHA-IIb (GPIIb) OR WITH ALPHA-V (VITRONECTIN RECEPTOR).			
CC	-1- PMM: THE CYSTEINE RESIDUES ARE INVOLVED IN INTRACHAIN DISULFIDE BONDS.			
CC	-1- SIMILARITY: BELONGS TO THE INTEGRIN BETA CHAIN FAMILY.			
CC	THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/or_send_an_email_to_license@isb-sib.ch">http://www.isb-sib.ch/announce/or_send_an_email_to_license@isb-sib.ch</a> ).			
CC	-----			
DR	EMBL; AF026509; AAB94086.1; -.			
DR	MGD; MGI:96612; Itgb3.			
DR	InterPro; IPR000561; -.			
DR	InterPro; IPR001169; -.			
DR	InterPro; IPR002369; -.			
DR	Pfam; PF00362; Integrin.B; 1.			
DR	PRINTS; PR01186; INTEGRIN.			
DR	PROSITE; PS00243; INTEGRIN_BETA_3.			
DR	PROSITE; PS00022; EGF_1; UNKNOWN_2.			
DR	PROSITE; PS01186; EGF_2; UNKNOWN_1.			
KW	Integrin; Cell adhesion; Transmembrane; Glycoprotein; Repeat;			
KW	Extracellular matrix; Cytoskeleton; Signal.			
FT	SIGNAL	1	25	POTENTIAL.
FT	CHAIN	26	787	PLATELET MEMBRANE GLYCOPROTEIN IIAA.
FT	DOMAIN	26	717	EXTRACELLULAR.
FT	TRANSMEM	718	740	POTENTIAL.
FT	DOMAIN	741	787	CYTOPLASMIC.
FT	DOMAIN	462	628	CYSTEINE-RICH REPEATS.
FT	REPEAT	462	510	I.
FT	REPEAT	511	552	II.
FT	REPEAT	553	591	III.
FT	REPEAT	592	628	IV.
FT	DISULFID	30	460	BY SIMILARITY.
FT	DISULFID	38	48	BY SIMILARITY.
FT	DISULFID	41	74	BY SIMILARITY.
FT	DISULFID	51	63	BY SIMILARITY.
FT	DISULFID	202	209	BY SIMILARITY.
FT	DISULFID	257	298	BY SIMILARITY.
FT	DISULFID	399	411	BY SIMILARITY.
FT	DISULFID	431	680	BY SIMILARITY.
FT	DISULFID	458	462	BY SIMILARITY.
FT	DISULFID	473	485	BY SIMILARITY.
FT	DISULFID	482	520	BY SIMILARITY.
FT	DISULFID	487	496	BY SIMILARITY.
FT	DISULFID	498	511	BY SIMILARITY.
FT	DISULFID	526	531	BY SIMILARITY.

FT DISULFID 528 561 BY SIMILARITY.  
 FT DISULFID 533 546 BY SIMILARITY.  
 FT DISULFID 548 553 BY SIMILARITY.  
 FT DISULFID 553 557 BY SIMILARITY.  
 FT DISULFID 567 572 BY SIMILARITY.  
 FT DISULFID 569 603 BY SIMILARITY.  
 FT DISULFID 574 583 BY SIMILARITY.  
 FT DISULFID 585 592 BY SIMILARITY.  
 FT DISULFID 606 611 BY SIMILARITY.  
 FT DISULFID 608 623 BY SIMILARITY.  
 FT DISULFID 613 623 BY SIMILARITY.  
 FT DISULFID 626 623 BY SIMILARITY.  
 FT DISULFID 633 643 BY SIMILARITY.  
 FT DISULFID 639 713 BY SIMILARITY.  
 FT DISULFID 660 681 BY SIMILARITY.  
 FT CARBOHYD 345 345 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 396 396 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 477 477 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 584 584 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 679 679 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 787 AA: 86694 MW: 81570599ABC438A3 CRC64:

Query Match 74.5%; Score 35; DB 1; Length 787;  
 Best Local Similarity 100.0%; Pred. No. 47;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 VCVOPG 8  
 DB 612 VCVOPG 617

RESULT 10  
 AMPM\_SALTY ID AMPM\_SALTY STANDARD; PRT; 263 AA.  
 AC P10882;  
 DT 01-JUL-1989 (Rel. 11, Created)  
 DT 01-NOV-1991 (Rel. 20, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE METHIONINE AMINOPEPTIDASE (EC 3.4.11.18) (MAP) (PEPTIDASE M).  
 GN MAP OR PEP.  
 OS Salmonella typhimurium.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Salmonella.  
 NX NCBI\_TaxID=602;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-LT2;  
 RX MEDLINE=91066851; PubMed=2250660;  
 RA Movva N.R., Semon D., Heyer C., Kawashima E., Wingfield P.,  
 Miller J.L., Miller C.G.;  
 RT "Cloning and nucleotide sequence of the Salmonella typhimurium pepm  
 gene."  
 RL Mol. Gen. Genet. 223:345-348(1990).  
 RN (2)  
 RP SEQUENCE.  
 RX MEDLINE=89210857; PubMed=2651123;  
 RA Wingfield P., Graber P., Turcatel G., Movva N.R., Pelletier M.,  
 Craig S., Rose K., Miller C.G.;  
 RT "Purification and characterization of a methionine-specific  
 aminopeptidase from Salmonella typhimurium."  
 RL Eur. J. Biochem. 180:23-32(1989).  
 CC -1- FUNCTION: REMOVES THE AMINO-TERMINAL METHIONINE FROM NASCENT  
 PROTEINS.  
 CC -1- CATALYTIC ACTIVITY: L-METHIONYLPEPTIDE + H(2)O = L-METHIONINE +  
 PEPTIDE.  
 CC -1- COFACTOR: COBALT.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M24A; ALSO KNOWN AS THE  
 MAP FAMILY 1.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: X55778; CA93298.1; -  
 DR PIR: S03562; S03562.  
 DR PIR: S12027; S12027.  
 DR HSSP: P07906; IMAT.  
 DR MEROPS: M24.001; -  
 DR StyGene: SG10283; map.  
 DR InterPro: IPR000994; -  
 DR InterPro: IPR001714; -  
 DR InterPro: IPR002467; -  
 DR Pfam: PF00557; Peptidase\_M24; 1.  
 DR PRINTS: PR00599; MAPEPTIDASE.  
 DR PROSITE: PS00680; MAP.1..1.  
 KW Hydroxylase; Aminopeptidase; Cobalt.  
 FT INT\_MET 0  
 FT METAL 96  
 FT METAL 107  
 FT METAL 170  
 FT METAL 203  
 FT METAL 235  
 FT CONFLICT 135  
 FT CONFLICT 144  
 FT CONFLICT 147  
 FT CONFLICT 158  
 SQ SEQUENCE 263 AA: 29160 MW: 88FEA56A36A8A78 CRC64:

Query Match 72.3%; Score 34; DB 1; Length 263;  
 Best Local Similarity 75.0%; Pred. No. 28;  
 Matches 6; Conservative 1; Mismatches 1; Indels 1; Gaps 0;

OY 1 GVCVOPG 8  
 DB 190 GVCVOPG 197

RESULT 11  
 STRE\_STRGR ID STRE\_STRGR STANDARD; PRT; 281 AA.  
 AC P09397;  
 DT 01-MAR-1989 (Rel. 10, Created)  
 DT 01-AUG-1992 (Rel. 23, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE STREPTOMYCIN BIOSYNTHESIS PROTEIN STRE.  
 GN STRE.  
 OS Streptomyces griseus.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.  
 NX NCBI\_TaxID=1911;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-N2-3-11;  
 RX MEDLINE=91375432; PubMed=1654502;  
 RA Mansouri K., Plepersberg W.;  
 RT "Genetics of streptomycin production in Streptomyces griseus:  
 nucleotide sequence of five genes, streghk, including a phosphatase  
 gene."  
 RL Mol. Gen. Genet. 228:459-469(1991).  
 CC -1- FUNCTION: MAY BE INVOLVED IN THE FORMATION OF N-METHYL-L-  
 GLUCOSAMINE.  
 CC -1- PATHWAY: STREPTOMYCIN BIOSYNTHESIS.  
 CC -1- SIMILARITY: LOW, TO THE AMINO-TERMINAL DOMAIN OF GLUCOSE-6-  
 PHOSPHATE ISOMERASE.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>



CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).  
CC -----  
DR EMBL; Y00459; CAA68518.1; -  
DR PIR; S17776; S17776.  
KM Streptomycin biosynthesis.  
SO SEQUENCE 281 AA; 31726 MW; 82456C5119B6387C CRC64;  
  
Query Match 72.3%; Score 34; DB 1; Length 281;  
Best Local Similarity 85.7%; Pred. No. 29;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
  
OY 2 GVCVOPG 8  
Db 5 GVCVOPG 11  
  
RESULT 12  
SHP1\_YEAST  
ID SHP1\_YEAST STANDARD; PRT; 423 AA.  
AC P34223;  
DT 01-FEB-1994 (Rel. 28, Created)  
DT 01-FEB-1994 (Rel. 28, Last sequence update)  
DT 01-NOV-1995 (Rel. 32, Last annotation update)  
DE SHP1 PROTEIN.  
GN SHP1 OR YBL058W OR YBL0509 OR YBL0515.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
OX NCBI\_TaxId=4932;  
[1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=5288C;  
RA Scherens B., el Bakoury M., Vierendeels F., Dubois E., Messenguy F.;  
RT "Sequencing and functional analysis of a 32,560 bp segment on the  
RT left arm of yeast chromosome II. Identification of 26 open reading  
RT frames, including the KIP1 and SEC17 genes.";  
RL Yeast 9.1355-1371(1993).  
RN  
RP IDENTIFICATION.  
RX MEDLINE=95198723; PubMed=7891699;  
RA Zhang S., Guha S., Volkert F.C.;  
RT "The Saccharomyces SHP1 gene, which encodes a regulator of  
RT phosphoprotein phosphatase 1 with differential effects on glycogen  
RT metabolism, meiotic differentiation, and mitotic cell cycle  
RT progression.";  
RL Mol. Cell. Biol. 15:2037-2050(1995).  
CC -1- FUNCTION: DIRECT OR INDIRECT POSITIVE REGULATOR OF GLC7 ACTIVITY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; 223261; CAA80789.1; -  
DR EMBL; 235819; CAA84878.1; -  
DR PIR; S39830; S39830.  
DR PIR; S37331; S37331.  
DR SGD; S0000154; SHP1.  
DR InterPro: IPR001012; -  
DR Pfam: PF00789; UBX; 1.  
SO SEQUENCE 423 AA; 46987 MW; AB46C8BD7ED4F11C CRC64;  
  
Query Match 72.3%; Score 34; DB 1; Length 423;  
Best Local Similarity 62.5%; Pred. No. 42;  
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
  
OY 1 GVCVOPG 8

Db 85 GGLCPKG 92  
||:|:|  
  
RESULT 13  
STHA\_MYCTU  
ID STHA\_MYCTU STANDARD; PRT; 468 AA.  
AC O0712;  
DT 01-OCT-2000 (Rel. 40, Created)  
DT 01-OCT-2000 (Rel. 40, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE PROBABLE SOLUBLE PYRIDINE NUCLEOTIDE TRANSHYDROGENASE (EC 1.6.1.1)  
DE (SMH) (NAD(P)(+) TRANSHYDROGENASE [B-SPECIFIC]).  
GN STHA OR RV2713 OR MYCT05A6.34.  
CN  
OS Mycobacterium tuberculosis.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
OX NCBI\_TaxId=1773;  
[1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=H37RV;  
RA MEDLINE=98295987; PubMed=9634230;  
RA Cole S.T., Broesch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E., III, Tekala F.,  
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
RA Davies R., Devlin K., Felkell T., Gentles S., Hamlin N., Holroyd S.,  
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,  
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
RA Rutter S., Seeger K., Skellon S., Squares S., Squares R., Sulston J.E.,  
RA Taylor K., Whitehead S., Barrett B.G.;  
RT "Deciphering the biology of Mycobacterium tuberculosis from the  
RT complete genome sequence.";  
RL Nature 393:537-544(1998).  
CC -1- FUNCTION: CONVERSION OF NADPH, GENERATED BY PERIPHERAL CATABOLIC  
CC PATHWAYS, TO NADH, WHICH CAN ENTER THE RESPIRATORY CHAIN FOR  
CC ENERGY GENERATION (BY SIMILARITY).  
CC -1- CATALYTIC ACTIVITY: NADPH + NAD(+) -> NADP(+) + NADH.  
CC -1- COFACTOR: FAD (BY SIMILARITY).  
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).  
CC -1- SIMILARITY: BELONGS TO THE PYRIDINE NUCLEOTIDE-DISULFIDE  
CC OXIDOREDUCTASES CLASS-I.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; 296072; CAB09473.1; -  
DR HSSP; P11959; IEBD.  
DR Tuberculin; RV2713; -  
DR InterPro: IPR000205; -  
DR InterPro: IPR001100; -  
DR Pfam; PF00070; Pyr\_redox; 1.  
DR PRINTS; PR00411; PNDROFAS1.  
DR PIR; PF00070; PYR\_redox; 1.  
KM Oxidoreductase; Flavoprotein; FAD; NADP; NAD.  
FT NP\_BIND 5 35 FAD (ADP PART) (BY SIMILARITY).  
FT NP\_BIND 300 310 FAD (FLAVIN PART) (BY SIMILARITY).  
SO SEQUENCE 468 AA; 50754 MW; D9E737C41C2898CF CRC64;  
  
Query Match 72.3%; Score 34; DB 1; Length 468;  
Best Local Similarity 75.0%; Pred. No. 45;  
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
  
OY 1 GVCVOPG 8  
Db 39 GVCVOPG 46  
  
RESULT 14

CAH\_DUNSA STANDARD: PRT: 589 AA.  
 ID CAH\_DUNSA  
 AC P54212;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 01-OCT-1996 (Rel. 34, Last annotation update)  
 DE CARBONIC ANHYDRASE (EC 4.2.1.1) (CARBONATE DEHYDRATASE).  
 GN DCA.  
 OS Dunalabella salina.  
 OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;  
 OC Dunalabiellaceae; Dunalabiella.  
 OX NCBI\_TaxID=3046;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=96279304; PubMed=8663366;  
 RA Fisher M., Gokhman I., Pick U., Zamir A.;  
 RT "A salt-resistant plasma membrane carbonic anhydrase is induced by  
 salt in Dunalabiella salina.";  
 RL J. Biol. Chem. 271:17718-17723(1996).  
 CC -1- FUNCTION: REVERSIBLE HYDRATATION OF CARBON DIOXIDE.  
 CC -1- CATALYTIC ACTIVITY: H(2)CO(3) -> CO(2) + H(2)O.  
 CC -1- INDUCTION: BY SALT.  
 CC -1- SIMILARITY: BELONGS TO THE EUKARYOTIC-TYPE CARBONIC ANHYDRASE  
 FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: U53811; AAC49378.1; -  
 DR HSSP: P00918; ICRA  
 DR InterPro: IPR001148; -  
 DR Pfam: PF00194; card\_anhydrase; 2.  
 DR Lyase; Zinc.  
 KM DOMAIN  
 FT METAL 390 589 CATALYTIC.  
 FT METAL 418 418 ZINC (CATALYTIC) (BY SIMILARITY).  
 FT METAL 420 420 ZINC (CATALYTIC) (BY SIMILARITY).  
 FT METAL 440 440 ZINC (CATALYTIC) (BY SIMILARITY).  
 SEQUENCE 589 AA; 64257 MW; 0CC2A6F42C121171 CRC64;  
 OY 1 GGCVCVOPG 8  
 DB 72 GGCVCVNTG 79  
 Query Match 72.3%; Score 34; DB 1; Length 589;  
 Best Local Similarity 75.0%; Pred. No. 55;  
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

RT cDNA clone. Identity with platelet glycoprotein IIa and similarity  
 RT to 'Integrin'.  
 RL J. Biol. Chem. 262:3936-3939(1987).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=90265363; PubMed=2345548;  
 RA Frachet P., Uzan G., Thevenon D., Denarier E., Prandini M.H.,  
 RA Marguerie G.;  
 RT GPIIb and GPIIa amino acid sequences deduced from human  
 RT megakaryocyte cDNAs.  
 RL Mol. Biol. Rep. 14:27-33(1990).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=88213696; PubMed=2452834;  
 RA Zimrin A.B., Elsmann R., Vilalre G., Schwartz E., Bennett J.S.,  
 RA Poncz M.;  
 RT "Structure of platelet glycoprotein IIa. A common subunit for two  
 RT different membrane receptors".  
 RL J. Clin. Invest. 81:1470-1475(1988).  
 RN [4]  
 RP SEQUENCE OF 27-788 FROM N.A.  
 RX MEDLINE=91009291; PubMed=2145280;  
 RA Lanza F., Kieffer N., Phillips D.R., Fitzgerald L.A.;  
 RT "Characterization of the human platelet glycoprotein IIa gene.  
 RT Comparison with the fibronectin receptor beta-subunit gene".  
 RL J. Biol. Chem. 265:18098-18103(1990).  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=90256778; PubMed=2341395;  
 RA Zimrin A.B., Gidwitz S., Lord S., Schwartz E., Bennett J.S.,  
 RA White G.C. II, Poncz M.;  
 RT "The genomic organization of platelet glycoprotein IIa".  
 RL J. Biol. Chem. 265:8590-8595(1990).  
 RN [6]  
 RP SEQUENCE OF 1-26 FROM N.A.  
 RC TISSUE-Blood;  
 RX MEDLINE=94129007; PubMed=8298129;  
 RA Villa-Garcia M., Li L., Riley G., Bray P.F.;  
 RT "Isolation and characterization of a TATA-less promoter for the human  
 RT beta 3 integrin gene".  
 RL Blood 83:668-676(1994).  
 RN [7]  
 RP SEQUENCE OF 122-204 FROM N.A.  
 RX MEDLINE=93002753; PubMed=1382574;  
 RA Jiang W.M., Jenkins D., Yuan Q., Leung E., Choo K.H., Watson J.D.,  
 RA Krissansen G.W.;  
 RT "The gene organization of the human beta 7 subunit, the common beta  
 RT subunit of the leukocyte integrins HML-1 and LPAW-1".  
 RL Int. Immunol. 4:1031-1040(1992).  
 RN [8]  
 RP SEQUENCE OF 218-234 AND 439-443.  
 RX MEDLINE=87101510; PubMed=3801670;  
 RA Hiraiwa A., Matsukage A., Shiku H., Takahashi T., Naito K., Yamada K.;  
 RT "Purification and partial amino acid sequence of human platelet  
 RT membrane glycoproteins IIb and IIa".  
 RL Blood 69:560-564(1987).  
 RN [9]  
 RP PARTIAL SEQUENCE, AND DISULFIDE BONDS.  
 RX MEDLINE=91158732; PubMed=2001252;  
 RA Calvele J.J., Henschen A., Gonzalez-Rodriguez J.;  
 RT "Assignment of disulphide bonds in human platelet GPIIa. A  
 RT disulphide pattern for the beta-subunits of the integrin family".  
 RL Biochem. J. 274:63-71(1991).  
 RN [10]  
 RP VARIANT HPA-1 (PL(A)).  
 RX MEDLINE=89214713; PubMed=2565345;  
 RA Newman P.J., Derbes R.S., Aster R.H.;  
 RT "The human platelet alloantigens, PL1 and PL2, are associated with  
 RT a leucine33/proline33 amino acid polymorphism in membrane  
 RT glycoprotein IIa, and are distinguishable by DNA typing".  
 RL J. Clin. Invest. 83:1778-1781(1989).  
 RN [11]  
 RP VARIANT HPA-4 (PEN).  
 RX MEDLINE=87165991; PubMed=3494014;  
 RA Fitzgerald L.A., Steiner B., Rall S.C. Jr., Lo S., Phillips D.R.;  
 RT "Protein sequence of endothelial glycoprotein IIa derived from a

RX MEDLINE-93055444; PubMed-1430225;  
 RA Wang R., Furihata K., McFarland J.G., Friedman K., Aster R.H.,  
 RA Newman P.J.;  
 RT "An amino acid polymorphism within the RGD binding domain of platelet  
 RT membrane glycoprotein IIb is responsible for the formation of the  
 RT Pensa/Penb alloantigen system.";  
 RL J. Clin. Invest. 90:2038-2043(1992).  
 RN [12]  
 RP VARIANT MO-1.  
 RX MEDLINE-93112977; PubMed-8093349;  
 RA Kullper R.W.A.M., Simsek S., Faber N.M., Goldschmeding R.,  
 RA van Weeren R.K.V., von Dem Borne A.E.G.K.;  
 RT "Single point mutation in human glycoprotein IIb is associated with  
 RT a new platelet-specific alloantigen (Mo) involved in neonatal  
 RT alloimmune thrombocytopenia.";  
 RL Blood 81:70-76(1993).  
 RN [13]  
 RP VARIANT CA/TU.  
 RX MEDLINE-94060373; PubMed-7694683;  
 RA Wang R., McFarland J.G., Kekomaki R., Newman P.J.;  
 RT "Amino acid 489 is encoded by a mutational 'hot spot' on the beta 3  
 RT integrin chain: the CA/TU human platelet alloantigen system.";  
 RL Blood 82:3386-3391(1993).  
 RN [14]  
 RP VARIANT SR(A).  
 RX MEDLINE-94179229; PubMed-8132570;  
 RA Santos S., Kald R., Kroll H., Walke M., Kiefel V.,  
 RA Mueller-Eckhardt C., Newman P.J.;  
 RT "A point mutation leads to an unpaired cysteine residue and a  
 RT molecular weight polymorphism of a functional platelet beta 3 integrin  
 RT subunit. The Sr.a alloantigen system of GPIIb.";  
 RL J. Biol. Chem. 269:8439-8444(1994).  
 RN [15]  
 RP VARIANT TYR-145.  
 RX MEDLINE-90364410; PubMed-2392682;  
 RA Loftus J.C., O'Toole T.E., Plov E.F., Glass A., Frelinger A.L. III,  
 RA Ginsberg M.H.;  
 RT "A beta 3 integrin mutation abolishes ligand binding and alters  
 RT divalent cation-dependent conformation.";  
 RL Science 249:915-918(1990).  
 RN [16]  
 RP VARIANT SER-240.  
 RX MEDLINE-92156115; PubMed-1371279;  
 RA Bajt M.L., Ginsberg M.H., Frelinger A.L. III, Berndt M.C.,  
 RA Loftus J.C.;  
 RT "A spontaneous mutation of integrin alpha IIb beta 3 (platelet  
 RT glycoprotein IIb-IIIa) helps define a ligand binding site.";  
 RL J. Biol. Chem. 267:3789-3794(1992).  
 RN [17]  
 RP VARIANT TRP-240.  
 RX MEDLINE-92291320; PubMed-1602006;  
 RA Lanza F., Stierle A., Fournier D., Morales M., Andre G., Nurdan A.T.,  
 RA Cazenave J.-P.;  
 RT "A new variant of Glanzmann's thrombasthenia (Strasbourg I).  
 RT Platelets with functionally defective glycoprotein IIb-IIIa complexes  
 RT and a glycoprotein IIa 214Arg-->214Trp mutation.";  
 RL J. Clin. Invest. 89:1995-2004(1992).  
 RN [18]  
 RP VARIANT STRASBOURG-1.  
 RX MEDLINE-93066201; PubMed-1438206;  
 RA Chen Y.-P., Daffar I., Pliard D., Steiner B., Cleutaut A.-M.,  
 RA Chen J.P., Rose J.-P.;  
 RT "Ser-752-->Pro mutation in the cytoplasmic domain of integrin beta 3  
 RT subunit and defective activation of platelet integrin alpha IIb beta  
 RT 3 (glycoprotein IIb-IIIa) in a variant of Glanzmann thrombasthenia.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 89:10169-10173(1992).  
 RN [19]  
 RP REVIEW ON GTA VARIANTS.  
 RX MEDLINE-95184171; PubMed-7878622;  
 RA Bray P.F.;  
 RT "Inherited diseases of platelet glycoproteins: considerations for  
 RT rapid molecular characterization.";  
 RL Thromb. Haemost. 72:492-502(1994).

RN [20]  
 RP VARIANTS GTA PRO-306; PHE-566 AND SER-598.  
 RX MEDLINE-99008899; PubMed-9790984;  
 RA Ando H., Kamata T., Handa M., Taki M., Kuwajima M., Kawai Y., Oda A.,  
 RA Murata M., Takada Y., Watanabe K., Ikeda Y.;  
 RT "Three novel integrin beta3 subunit missense mutations (H280P, C560F,  
 RT and G579S) in thrombasthenia, including one (H280P) prevalent in  
 RT Japanese patients.";  
 RL Biochem. Biophys. Res. Commun. 251:763-768(1998).  
 CC -I- FUNCTION: PLATELET MEMBRANE GLYCOPROTEIN IIb-IIIa IS THE PLATELET  
 CC RECEPTOR FOR ADHESIVE PROTEINS. IT BINDS FIBRINOGEN, FIBRONECTIN,  
 CC VITRONECTIN AND VWF. INTEGRINS RECOGNIZE THE SEQUENCE R-G-D IN  
 CC THEIR LIGAND.  
 CC -I- SUBUNIT: DIMER OF AN ALPHA AND BETA SUBUNIT. BETA-3 IS EITHER  
 CC ASSOCIATED WITH ALPHA-IIb (GPIIb) OR WITH ALPHA-V (VITRONECTIN  
 CC RECEPTOR).  
 CC -I- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.  
 CC -I- ALTERNATIVE PRODUCTS: THE DIFFERENT FORMS OF BETA-3 SUBUNIT MAY  
 CC ARISE BY ALTERNATIVE SPLICING OF PRIMARY MRNA TRANSCRIPTS.  
 CC -I- PTM: THE CYSTEINE RESIDUES ARE INVOLVED IN INTRACHAIN DISULFIDE  
 CC BONDS.  
 CC -I- POLYMORPHISM: POSITION 59 IS ASSOCIATED WITH PLATELET-SPECIFIC  
 CC ALLOANTIGEN HPA-1 (2W OR PL(A)). HPA-1A/PL(A1) HAS LEU-59 AND  
 CC HPA-1B/PL(A2) HAS PRO-59.  
 CC -I- POLYMORPHISM: POSITION 169 IS ASSOCIATED WITH PLATELET-SPECIFIC  
 CC ALLOANTIGEN HPA-4 (PEN OR YUK). HPA-4A/PEN(A)/YUK(A) HAS ARG-169  
 CC AND HPA-4B/PEN(B)/YUK(B) HAS GLN-169. HPA-4B IS INVOLVED IN  
 CC NEONATAL ALLOIMMUNE THROMBOCYTOPENIA (NATP).  
 CC -I- POLYMORPHISM: POSITION 433 IS ASSOCIATED WITH PLATELET-SPECIFIC  
 CC ALLOANTIGEN MO. MO(-) HAS PRO-433 AND MO(+) HAS ALA-433. MO(+) IS  
 CC INVOLVED IN NATP.  
 CC -I- POLYMORPHISM: POSITION 515 IS ASSOCIATED WITH PLATELET-SPECIFIC  
 CC ALLOANTIGEN CA (T0). CA(-)/TU(-) HAS ARG-515 AND CA(1)/TU(+) HAS  
 CC GLN-515. CA(+) IS INVOLVED IN NATP.  
 CC -I- DISEASE: DEFECTS IN ITGB3 ARE ONE OF THE CAUSE OF GLANZMANN  
 CC THROMBASTHENIA (GTA), AN AUTOSOMAL RECESSIVE DISORDER WHICH IS  
 CC THE MOST COMMON INHERITED DISEASE OF PLATELETS. GTA IS

Query Match 72.3%; Score 34; DB 1; Length 788;  
 Best Local Similarity 83.3%; Pred. No. 71;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy 3 VCVOPG 8  
 Db 613 VCIOPG 618

Search completed: June 13, 2001, 14:21:46  
 Job time: 806 sec

Wed Jun 13 14:59:41 2001

pct-us01-05825a-13.rsp

---

Page 10



```

ID 075851 PRELIMINARY: PRT; 4123 AA.
AC 075851;
DT 01-NOV-1998 (TREMblrel. 08, Created)
DT 01-NOV-1998 (TREMblrel. 08, Last sequence update)
DE 01-OCT-2000 (TREMblrel. 15, Last annotation update)
DE WUGSC:H.DJ0751H13.1 PROTEIN (FRAGMENT).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN 1;
RP SEQUENCE FROM N.A.
RA Leonard S., Graves T., Strommatt C.;
RT "The sequence of Homo sapiens PAC clone DJ0751H13."
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
RN 12;
RP SEQUENCE FROM N.A.
RA Waterston R.H.;
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
RN 13;
RP SEQUENCE FROM N.A.
RA Waterston R.;
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC004877; AAC3631.1; -
DR HSSP; P01130; IATJ
DR INTERPRO; IPR000421; -
DR INTERPRO; IPR000561; -
DR INTERPRO; IPR000884; -
DR INTERPRO; IPR000923; -
DR INTERPRO; IPR001007; -
DR INTERPRO; IPR001064; -
DR INTERPRO; IPR001092; -
DR INTERPRO; IPR001846; -
DR INTERPRO; IPR002172; -
DR INTERPRO; IPR002223; -
DR INTERPRO; IPR002465; -
DR INTERPRO; IPR002919; -
DR PFAM; PF00057; 1d1_recept_a; 11.
DR PFAM; PF00090; tsp_1; 14.
DR PFAM; PF00094; wtd_5; 5.
DR PFAM; PF00754; p5_F8_type_C; 1.
DR PRAM; PF01826; TIL; 9.
DR PRINTS; PR00261; LDLRECEPTOR.
DR PROSITE; PS00022; EGF_1; UNKNOWN_1.
DR PROSITE; PS00196; COPPER_BLUE; UNKNOWN_1.
DR PROSITE; PS00225; CRYSTALLIN_BETAGAMMA; UNKNOWN_1.
DR PROSITE; PS00280; BPT_KUNITZ; UNKNOWN_1.
DR PROSITE; PS00340; RECEPTOR_CYTOKINES_2; UNKNOWN_6.
DR PROSITE; PS01209; LDLRA_1; 9.
DR PROSITE; PS50068; LDLRA_2; 20.
KM Glycoprotein.
FN NON_TER
SQ SEQUENCE 4123 AA; 434985 MW; 7AAB6F8BDCE012PB CRC64;

```

```

Query Match 85.1%; Score 40; DB 4; Length 4123;
Best Local Similarity 75.0%; Pred. No. 76;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```

```

QY 1 GGCYVOP 8
DB 3234 GATCYVOP 3241

```

```

RESULT 3
P95220 PRELIMINARY: PRT; 300 AA.
AC P95220;
DT 01-MAY-1997 (TREMblrel. 03, Created)
DT 01-MAY-1997 (TREMblrel. 03, Last sequence update)
DE 01-JUN-2000 (TREMblrel. 14, Last annotation update)
DE HYPOTHEICAL 32.2 KDA PROTEIN.
GN RV0263C OR MTCY06A4.07C.

```

```

OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN 1;
RP SEQUENCE FROM N.A.
RA STRAIN-H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hovner S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squires R., Sulston J.E.,
RA Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544(1998).
DR EMBL; Z86089; CAB06590.1; -
DR TUBERCULIST; RV0263C; -
KW Hypothetical protein.
SQ SEQUENCE 300 AA; 32203 MW; 98E4350A8949F8D4 CRC64;

```

```

Query Match 80.9%; Score 38; DB 2; Length 300;
Best Local Similarity 85.7%; Pred. No. 15;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 GGCYVOP 7
DB 120 GGCYVOP 126

```

```

RESULT 4
Q9W6F8 PRELIMINARY: PRT; 374 AA.
AC Q9W6F8;
DT 01-NOV-1999 (TREMblrel. 12, Created)
DT 01-NOV-1999 (TREMblrel. 12, Last sequence update)
DE 01-OCT-2000 (TREMblrel. 15, Last annotation update)
DE WNT INHIBITORY FACTOR 1 PRECURSOR (WNT INHIBITORY FACTOR-1) (WIF-1).
GN WIF1 OR WIF-1.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipridae; Pipridae;
OX NCBI_TaxID=8355;
RN 1;
RP SEQUENCE FROM N.A.
RX MEDLINE=99215557; PubMed=10201374;
RA Hsieh J.-C., Kodjabachian L., Rebert M.L., Ratner A.,
RA Smallwood P.M., Samos C.H., Nusse R., David I.B., Nathans J.;
RT "A new secreted protein that binds to Wnt proteins and inhibits their
RT activities."
RL Nature 398:431-436(1999).
CC -1- FUNCTION: BINDS TO WNT PROTEINS AND INHIBITS THEIR ACTIVITIES. MAY
CC BE INVOLVED IN MESODERM SEGMENTATION.
CC -1- TISSUE SPECIFICITY: DURING SOMATOGENESIS, EXPRESSED PREDOMINANTLY
CC IN UNSEGMENTED PARAXIAL PRESOMITIC MESODERM AND, TO A MUCH LESSER
CC EXTENT, IN NEWLY SEGMENTED SOMITES.
CC -1- DEVELOPMENTAL STAGE: FIRST EXPRESSED AT NEURULA STAGES.
CC EMBL; AF122924; AAD25404.1; -
DR HSSP; P00740; IEDM.
DR INTERPRO; IPR000561; -
DR INTERPRO; IPR002049; -
DR PFAM; PF00008; EGF_5.
DR PRINTS; PR00011; EGFAMININ.
DR PROSITE; PS00022; EGF_1; UNKNOWN_5.
DR PROSITE; PS01186; EGF_2; 5.
KW Repeat; EGF-like domain; Signal; Developmental protein.
FT SIGNAL 1
FT 28
FT POTENTIAL.

```

FT CHAIN 29 374 WNT INHIBITORY FACTOR 1.  
 FT DOMAIN 172 203 EGF-LIKE 1.  
 FT DOMAIN 204 235 EGF-LIKE 2.  
 FT DOMAIN 236 267 EGF-LIKE 3.  
 FT DOMAIN 268 299 EGF-LIKE 4.  
 FT DOMAIN 300 331 EGF-LIKE 5.  
 FT DISULFID 172 181 POTENTIAL.  
 FT DISULFID 177 187 POTENTIAL.  
 FT DISULFID 193 195 POTENTIAL.  
 FT DISULFID 204 213 POTENTIAL.  
 FT DISULFID 209 219 POTENTIAL.  
 FT DISULFID 225 227 POTENTIAL.  
 FT DISULFID 236 245 POTENTIAL.  
 FT DISULFID 241 251 POTENTIAL.  
 FT DISULFID 257 259 POTENTIAL.  
 FT DISULFID 268 277 POTENTIAL.  
 FT DISULFID 273 283 POTENTIAL.  
 FT DISULFID 289 291 POTENTIAL.  
 FT DISULFID 300 309 POTENTIAL.  
 FT DISULFID 305 315 POTENTIAL.  
 FT DISULFID 321 323 POTENTIAL.  
 SQ SEQUENCE 374 AA; 41071 MW; E26F973B0F00ACF8 CRC64;

Query Match Best Local Similarity 80.9%; Score 38; DB 13; Length 374;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGCVCVQPG 8  
 DB 216 GGLCVTPG 223

RESULT 5  
 Q9Y5W5 PRELIMINARY; PRT; 379 AA.  
 AC Q9Y5W5;  
 DT 01-NOV-1999 (TREMBLERel. 12, Created)  
 DT 01-NOV-1999 (TREMBLERel. 12, Last sequence update)  
 DT 01-OCT-2000 (TREMBLERel. 15, Last annotation update)  
 DE WNT INHIBITORY FACTOR 1 PRECURSOR (WNT INHIBITORY FACTOR-1) (WIF-1).  
 GN WIF1 OR WIF-1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 OC NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99215557; PubMed=10201374;  
 RA Hsieh J.-C., Kodjabachian L., Rebber M.L., Rattner A.,  
 RA Smallwood P.M., Samos C.H., Nusse R., David I.B., Nathans J.;  
 RT "A new secreted protein that binds to Wnt proteins and inhibits their  
 RT activities.";  
 RL Nature 398:431-436(1999).  
 CC -1- FUNCTION: BINDS TO WNT PROTEINS AND INHIBITS THEIR ACTIVITIES. MAY  
 CC BE INVOLVED IN MESODERM SEGMENTATION.  
 CC -1- SIMILARITY: CONTAINS 5 EGF-LIKE DOMAINS.  
 CC EMBL; AF122922; AAD25402.1; -.  
 DR HSSP; P00743; ICGF.  
 DR INTERPRO; IPR000561; -.  
 DR INTERPRO; IPR002049; -.  
 DR PFAM; PF00008; EGF\_5.  
 DR PRINTS; PR00011; EGF\_LAMININ.  
 DR PROSITE; PS00022; EGF\_1; UNKNOWN\_5.  
 DR Repeat; EGF-like domain; Signal; Developmental protein.  
 KW Repeat; EGF-like domain; Signal; Developmental protein.  
 FT CHAIN 29 379 WNT INHIBITORY FACTOR 1.  
 FT SIGNAL 1 28 POTENTIAL.  
 FT DOMAIN 177 208 EGF-LIKE 1.  
 FT DOMAIN 209 240 EGF-LIKE 2.  
 FT DOMAIN 241 272 EGF-LIKE 3.  
 FT DOMAIN 273 304 EGF-LIKE 4.  
 FT DOMAIN 305 336 EGF-LIKE 5.  
 FT DISULFID 321 323 POTENTIAL.

FT DISULFID 177 186 POTENTIAL.  
 FT DISULFID 182 192 POTENTIAL.  
 FT DISULFID 198 200 POTENTIAL.  
 FT DISULFID 209 218 POTENTIAL.  
 FT DISULFID 214 224 POTENTIAL.  
 FT DISULFID 230 232 POTENTIAL.  
 FT DISULFID 241 250 POTENTIAL.  
 FT DISULFID 246 256 POTENTIAL.  
 FT DISULFID 262 264 POTENTIAL.  
 FT DISULFID 273 282 POTENTIAL.  
 FT DISULFID 278 288 POTENTIAL.  
 FT DISULFID 294 296 POTENTIAL.  
 FT DISULFID 305 314 POTENTIAL.  
 FT DISULFID 310 320 POTENTIAL.  
 FT DISULFID 326 328 POTENTIAL.  
 SQ SEQUENCE 379 AA; 41512 MW; 27782370A266E784 CRC64;

Query Match Best Local Similarity 80.9%; Score 38; DB 4; Length 379;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGCVCVQPG 8  
 DB 221 GGLCVTPG 228

RESULT 6  
 Q9WUJ1 PRELIMINARY; PRT; 379 AA.  
 AC Q9WUJ1;  
 DT 01-NOV-1999 (TREMBLERel. 12, Created)  
 DT 01-NOV-1999 (TREMBLERel. 12, Last sequence update)  
 DT 01-OCT-2000 (TREMBLERel. 15, Last annotation update)  
 DE WNT INHIBITORY FACTOR 1 PRECURSOR (WNT INHIBITORY FACTOR-1) (WIF-1).  
 GN WIF1 OR WIF-1.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OC NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99215557; PubMed=10201374;  
 RA Hsieh J.-C., Kodjabachian L., Rebber M.L., Rattner A.,  
 RA Smallwood P.M., Samos C.H., Nusse R., David I.B., Nathans J.;  
 RT "A new secreted protein that binds to Wnt proteins and inhibits their  
 RT activities.";  
 RL Nature 398:431-436(1999).  
 CC -1- FUNCTION: BINDS TO WNT PROTEINS AND INHIBITS THEIR ACTIVITIES. MAY  
 CC BE INVOLVED IN MESODERM SEGMENTATION.  
 CC -1- TISSUE SPECIFICITY: EXPRESSION HIGHEST IN HEART AND LUNG. LOWER IN  
 CC BRAIN AND EYE.  
 CC -1- SIMILARITY: CONTAINS 5 EGF-LIKE DOMAINS.  
 CC EMBL; AF122923; AAD25403.1; -.  
 DR HSSP; P00740; IIXA.  
 DR MGD; MGI:1344332; Wif1-pending.  
 DR INTERPRO; IPR000561; -.  
 DR INTERPRO; IPR002049; -.  
 DR PFAM; PF00008; EGF\_5.  
 DR PRINTS; PR00011; EGF\_LAMININ.  
 DR PROSITE; PS00022; EGF\_1; UNKNOWN\_5.  
 DR Repeat; EGF-like domain; Signal; Developmental protein.  
 KW Repeat; EGF-like domain; Signal; Developmental protein.  
 FT CHAIN 29 379 WNT INHIBITORY FACTOR 1.  
 FT SIGNAL 1 28 POTENTIAL.  
 FT DOMAIN 177 208 EGF-LIKE 1.  
 FT DOMAIN 209 240 EGF-LIKE 2.  
 FT DOMAIN 241 272 EGF-LIKE 3.  
 FT DOMAIN 273 304 EGF-LIKE 4.  
 FT DOMAIN 305 336 EGF-LIKE 5.  
 FT DISULFID 177 186 POTENTIAL.  
 FT DISULFID 182 192 POTENTIAL.  
 FT DISULFID 198 200 POTENTIAL.

FT DISULFID 209 216 POTENTIAL.  
 FT DISULFID 214 224 POTENTIAL.  
 FT DISULFID 230 232 POTENTIAL.  
 FT DISULFID 241 250 POTENTIAL.  
 FT DISULFID 246 256 POTENTIAL.  
 FT DISULFID 262 264 POTENTIAL.  
 FT DISULFID 273 282 POTENTIAL.  
 FT DISULFID 278 288 POTENTIAL.  
 FT DISULFID 294 296 POTENTIAL.  
 FT DISULFID 305 314 POTENTIAL.  
 FT DISULFID 310 320 POTENTIAL.  
 FT DISULFID 326 328 POTENTIAL.  
 SQ SEQUENCE 379 AA; 41590 MW; E3765F2642B2BC9A CRC64;

Query Match  
 Best Local Similarity 80.9%; Score 38; DB 11; Length 379;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 GGVCPQPG 8  
 DB 221 GGLCVTPG 228

RESULT 7  
 ID 09UH32 PRELIMINARY; PRT; 531 AA.  
 AC 09UH32; 01-MAY-2000 (TREMBlrel. 13, Created)  
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
 DT 01-JUN-2000 (TREMBlrel. 14, Last annotation update)  
 DE D1104E15.1 (MANNOsyl. (BETA-1,4)-GLYCOPROTEIN  
 DE BETA-1,4-N-ACETYLGLUCOSAMINYLTRANSFERASE).  
 GN MGAT3.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Barlow K.;  
 RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AL022312; CAB45282.1; -;  
 DR INTERPRO: IPR000561; -;  
 DR PROSITE: PS00022; EGF\_1; UNKNOWN\_1.  
 DR PROSITE: PS01186; EGF\_2; 1.  
 KW transferase; Glycosyltransferase.  
 SQ SEQUENCE 531 AA; 61053 MW; 3CF56C518F6248B0 CRC64;

Query Match  
 Best Local Similarity 80.9%; Score 38; DB 4; Length 531;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 GGVCPQPG 8  
 DB 112 GGVCPKPG 119

RESULT 8  
 ID 075413 PRELIMINARY; PRT; 179 AA.  
 AC 075413; 01-NOV-1998 (TREMBlrel. 08, Created)  
 DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)  
 DT 01-MAY-2000 (TREMBlrel. 13, Last annotation update)  
 DE LATENT TRANSFORMING GROWTH FACTOR-BETA BINDING PROTEIN 4L (FRAGMENT).  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.

RC TISSUE-HEART;  
 RX MEDLINE=98325059; PubMed=9660815;  
 RA Saharinen J., Taipale J., Monni O., Keski-Oja J.;  
 RT "Identification and characterization of a new latent transforming  
 RT growth factor-beta binding protein, LTBP-4.";  
 RL J. Biol. Chem. 273:18459-18469(1998).  
 DR EMBL: AF051345; AAC39880.1; -;  
 DR INTERPRO: IPR000561; -;  
 DR PFM: PF00008; EGF\_1.  
 DR PROSITE: PS00022; EGF\_1; UNKNOWN\_1.  
 FT NON\_TER 1  
 FT NON\_TER 179  
 SQ SEQUENCE 179 AA; 19062 MW; 137530C3FC8B2131 CRC64;

Query Match  
 Best Local Similarity 78.7%; Score 37; DB 4; Length 179;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 GGVCPQPG 7  
 DB 140 GGVCPKPG 146

RESULT 9  
 ID 099944 PRELIMINARY; PRT; 293 AA.  
 AC 099944; 01-MAY-1997 (TREMBlrel. 03, Created)  
 DT 01-MAY-1997 (TREMBlrel. 03, Last sequence update)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)  
 DE HYPOTHETICAL 32.3 KDA PROTEIN.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Rowen L., Dankers C., Baskin D., Faust J., Loretz C., Ahearn M.E.,  
 RA Banta A., Spies T., Hood L.;  
 RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: U89335; AAB47494.1; -;  
 DR HSSP: P35555; IEMN.  
 DR INTERPRO: IPR000152; -;  
 DR INTERPRO: IPR000561; -;  
 DR INTERPRO: IPR001881; -;  
 DR PFM: PF00008; EGF\_2.  
 DR PROSITE: PS00010; ASX\_HYDROXYL; 1.  
 DR PROSITE: PS00022; EGF\_1; UNKNOWN\_1.  
 DR PROSITE: PS01186; EGF\_2; 1.  
 DR PROSITE: PS01187; EGF\_CA; 1.  
 KW Hypothetical protein; Glycoprotein; EGF-like domain.  
 SQ SEQUENCE 293 AA; 32262 MW; 6519CA254FB23FD0 CRC64;

Query Match  
 Best Local Similarity 78.7%; Score 37; DB 4; Length 293;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 GGVCPQPG 7  
 DB 121 GGVCPKPG 127

RESULT 10  
 ID 094106 PRELIMINARY; PRT; 645 AA.  
 AC 094106; 01-MAY-1999 (TREMBlrel. 10, Created)  
 DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)  
 DE HEAT SHOCK PROTEIN 70 (FRAGMENT).  
 GN PCSAL.



```
OS Pneumocystis carinii f. sp. carinii.
OC Eukaryota; Fungi; Fungi Incertae sedis; Pneumocystidaceae;
XX Pneumocystis.
XX NCBI_TaxID=38081;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97045128; PubMed=8890193;
RA Stedman T.T.; Buck G.A.;
RT "Identification, characterization, and expression of the BLP
  endoplasmic reticulum resident chaperonins in Pneumocystis carinii.";
RL Infect. Immun. 64:4463-4471(1996).
DR EMBL; U80967; AAD00455.1; -.
DR HSSP; P19120; 3HSC.
DR INTERPRO; IPR000169; -.
DR INTERPRO; IPR001023; -.
DR PFAM; PF00012; HSP70; 1.
DR PRINTS; PR00301; HEATSHOCK70.
DR PROSITE; PS00297; HSP70_1; UNKNOWN_1.
DR PROSITE; PS00329; HSP70_2; 1.
DR PROSITE; PS00639; THIOL_PROTEASE_HIS; UNKNOWN_1.
DR PROSITE; PS01036; HSP70_3; 1.
KM Heat shock.
FT NON_TER
SQ SEQUENCE 645 AA; 70884 MM; 6D8CF90433BB163F CRC64;

Query Match      78.7%; Score 37; DB 3; Length 645;
Best Local Similarity 75.0%; Pred. NO. 48;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 GGCVCVPC 8
DB 621 GGACSQPC 628

RESULT 11
ID 094104 PRELIMINARY; PRT: 647 AA.
AC 094104;
DT 01-MAY-1999 (TREMBlrel. 10, Created)
DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
DE HEAT SHOCK PROTEIN 70 (FRAGMENT).
GN HSP70.
OS Pneumocystis carinii.
OC Eukaryota; Fungi; Fungi Incertae sedis; Pneumocystidaceae;
XX Pneumocystis.
XX NCBI_TaxID=4754;
RN [1]
RP SEQUENCE FROM N.A.
RA Paul S.P.; Graves D.C.;
RT "Phylogeny and sequence analysis of Pneumocystis carinii HSP70.";
RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; U71151; AAD09565.1; -.
DR HSSP; P19120; 3HSC.
DR INTERPRO; IPR000169; -.
DR INTERPRO; IPR001023; -.
DR PFAM; PF00012; HSP70; 1.
DR PRINTS; PR00301; HEATSHOCK70.
DR PROSITE; PS00297; HSP70_1; UNKNOWN_1.
DR PROSITE; PS00329; HSP70_2; 1.
DR PROSITE; PS00639; THIOL_PROTEASE_HIS; UNKNOWN_1.
DR PROSITE; PS01036; HSP70_3; 1.
KM Heat shock.
FT NON_TER
SQ SEQUENCE 647 AA; 71176 MM; 92F9A96399380F9 CRC64;

Query Match      78.7%; Score 37; DB 3; Length 647;
Best Local Similarity 75.0%; Pred. NO. 48;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 GGCVCVPC 8
DB 621 GGACSQPC 628
```

```
DB 623 GGACSQPC 630

RESULT 12
ID 09WU59 PRELIMINARY; PRT: 700 AA.
AC 09WU59;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-MAY-2000 (TREMBlrel. 13, Last annotation update)
DE HEDGEHOG-INTERACTING PROTEIN.
GN HIP.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
XX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99158142; PubMed=10050855;
RA Chang P.T.; McMahon A.P.;
RT "Vertebrate Hedgehog signalling modulated by induction of a Hedgehog-
  binding protein.";
RL Nature 397:617-621(1999).
DR EMBL; AF116865; AAD31172.1; -.
DR INTERPRO; IPR000561; -.
DR PROSITE; PS00022; EGF_1; UNKNOWN_2.
DR PROSITE; PS01186; EGF_2; 2.
SQ SEQUENCE 700 AA; 78485 MM; DC19B3929AFC6354 CRC64;

Query Match      78.7%; Score 37; DB 11; Length 700;
Best Local Similarity 85.7%; Pred. NO. 52;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GGCVCVPC 7
DB 646 GGVCVPC 652

RESULT 13
ID 002660 PRELIMINARY; PRT: 870 AA.
AC 002660;
DT 01-JUL-1997 (TREMBlrel. 04, Created)
DT 01-JAN-1999 (TREMBlrel. 09, Last sequence update)
DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
DE SCO-SPONDIN (FRAGMENT).
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
XX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-SUBCOMMISSURAL ORGAN;
RA Gobron S.;
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; Y08560; CAA69867.1; -.
DR HSSP; P01130; 1A1J
DR INTERPRO; IPR000884; -.
DR INTERPRO; IPR001846; -.
DR INTERPRO; IPR002172; -.
DR INTERPRO; IPR002465; -.
DR INTERPRO; IPR002919; -.
DR PFAM; PF00057; 1d1_recept_a; 7.
DR PFAM; PF00090; tsp_1; 1.
DR PFAM; PF00094; vwd; 1.
DR PFAM; PF01826; TIL; 2.
DR PRINTS; PR00261; LDLRECEPTOR.
DR PROSITE; PS00340; RECEPTOR_CYOKINES_2; UNKNOWN_2.
DR PROSITE; PS01209; LDLRA_1; 4.
DR PROSITE; PS50068; LDLRA_2; 6.
```

KW Glycoprotein. 1  
 FT NON\_TER 870 87C  
 FT NON\_TER 870 87C  
 SQ SEQUENCE 870 AA; 91875 MW; DD9DAF7B77AC48CE CRC64;

Query Match  
 Best Local Similarity 78.7%; Score 37; DB 6; Length 870;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 OY 1 GGVCVOPG 8  
 DB 474 GGCVPEPG 481

RESULT 14  
 075412  
 ID 075412 PRELIMINARY; PRT; 1511 AA.  
 AC 075412:  
 DT 01-NOV-1998 (TREMBlrel. 08, Created)  
 DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)  
 DE LATENT TRANSFORMING GROWTH FACTOR-BETA BINDING PROTEIN 4S.  
 OS Homo sapiens (Human).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=HEART;  
 RX MEDLINE=98325059; PubMed=9660815;  
 RA Saharinen J., Taipale J., Monni O., Keski-Oja J.;  
 RT "Identification and characterization of a new latent transforming  
 growth factor-beta-binding protein, LTBP-4.";  
 RL J Biol. Chem. 273:18439-18469(1998).  
 DR EMBL: AF051344; AAC39879.1; -;  
 DR HSSP: P35555; IEMN.  
 DR INTERPRO: IPR000152; -;  
 DR INTERPRO: IPR000561; -;  
 DR INTERPRO: IPR001881; -;  
 DR INTERPRO: IPR002212; -;  
 DR PRAM: PR00008; EGF\_17.  
 DR PRAM: PR00683; TB; 4.  
 DR PROSITE: PS00010; ASX\_HYDROXYL; 14.  
 DR PROSITE: PS00022; EGF\_1; UNKNOWN\_2.  
 DR PROSITE: PS01186; EGF\_2; 12.  
 DR PROSITE: PS01187; EGF\_CA; 17.  
 KW Glycoprotein; EGF-like domain.  
 SQ SEQUENCE 1511 AA; 161157 MW; C61AB757B256958D CRC64;

Query Match  
 Best Local Similarity 78.7%; Score 37; DB 4; Length 1511;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 OY 1 GGVCVOP 7  
 DB 47 GGVCVOP 53

RESULT 15  
 000508  
 ID 000508 PRELIMINARY; PRT; 1587 AA.  
 AC 000508:  
 DT 01-JUL-1997 (TREMBlrel. 04, Created)  
 DT 01-JUL-1997 (TREMBlrel. 04, Last sequence update)  
 DE LATENT TGF-BETA BINDING PROTEIN-4.  
 OS Homo sapiens (Human).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]

RP SEQUENCE FROM N.A.  
 RX MEDLINE=97415399; PubMed=9271198;  
 RA Giltay R., Kostka G., Timpi R.;  
 RT "Sequence and expression of a novel member (LTBP-4) of the family of  
 latent transforming growth factor-beta binding proteins.";  
 RL FEBS Lett. 411:164-168(1997).  
 DR EMBL: Y13622; CAJ73944.1; -;  
 DR HSSP: P35555; IEMN.  
 DR INTERPRO: IPR000152; -;  
 DR INTERPRO: IPR000561; -;  
 DR INTERPRO: IPR001881; -;  
 DR INTERPRO: IPR002212; -;  
 DR PRAM: PR00008; EGF\_17.  
 DR PRAM: PR00683; TB; 4.  
 DR PROSITE: PS00010; ASX\_HYDROXYL; 14.  
 DR PROSITE: PS00022; EGF\_1; UNKNOWN\_2.  
 DR PROSITE: PS01186; EGF\_2; 12.  
 DR PROSITE: PS01187; EGF\_CA; 17.  
 KW Glycoprotein; EGF-like domain.  
 SQ SEQUENCE 1587 AA; 169620 MW; 57A832F95FA0AE46 CRC64;

Query Match  
 Best Local Similarity 78.7%; Score 37; DB 4; Length 1587;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 OY 1 GGVCVOP 7  
 DB 123 GGVCVOP 129

Search completed: June 13, 2001, 14:20:33  
 Job time: 733 sec



*This Page Blank (uspto)*

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

## OM protein - protein search, using sw model

Run on: June 13, 2001, 14:16:35 ; Search time 118.55 Seconds  
(without alignments)  
1.296 Million cell updates/sec

Title: PCT-US01-05825a-13

Perfect score: 47

Sequence: 1 GGYCVDPG 8

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 185757 seqs, 19210857 residues

Total number of hits satisfying chosen parameters: 185757

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA.\*  
1: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep.\*  
2: /cgn2\_6/ptodata/2/1aa/5B.COMB.pep.\*  
3: /cgn2\_6/ptodata/2/1aa/6A.COMB.pep.\*  
4: /cgn2\_6/ptodata/2/1aa/6B.COMB.pep.\*  
5: /cgn2\_6/ptodata/2/1aa/PCTRUS.COMB.pep.\*  
6: /cgn2\_6/ptodata/2/1aa/backfile1.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	38	80.9	531	2	US-08-975-114A-4
2	38	80.9	531	4	US-08-849-281A-4
3	38	80.9	536	2	US-08-975-114A-5
4	38	80.9	536	4	US-08-849-281A-3
5	37	78.7	20	2	US-08-859-931A-4
6	37	78.7	118	3	US-08-545-809A-125
7	34	72.3	92	2	US-07-728-215-39
8	34	72.3	92	2	US-07-728-215-41
9	34	72.3	718	1	US-08-444-792-4
10	34	72.3	788	2	US-07-728-215-32
11	34	72.3	788	1	US-08-444-792-4
12	33	70.2	118	3	US-08-545-809A-108
13	33	70.2	119	2	US-08-475-000-16
14	33	70.2	119	2	US-08-483-199-16
15	33	70.2	119	2	US-08-484-508-16
16	33	70.2	123	2	US-08-428-197-38
17	33	70.2	123	2	US-08-428-197-38
18	33	70.2	123	5	PCT-US93-10555-38
19	33	70.2	1239	5	US-08-937-931-2
20	32	68.1	120	1	US-09-285-502-2
21	32	68.1	120	1	US-08-478-039-101
22	32	68.1	573	2	US-08-476-349A-101
23	32	68.1	573	2	US-08-745-934-4
24	32	68.1	573	2	US-09-147-009-7
25	31	66.0	770	1	US-08-445-135-2
26	31	66.0	130	4	US-09-347-819-8
27	31	66.0	181	1	US-08-278-091-15

28	31	66.0	181	1	US-08-483-859-15	Sequence 15, Appl
29	31	66.0	181	1	US-08-472-173-15	Sequence 15, Appl
30	31	66.0	181	2	US-08-487-167-15	Sequence 15, Appl
31	31	66.0	181	2	US-08-482-816-15	Sequence 15, Appl
32	31	66.0	181	2	US-08-296-149-15	Sequence 15, Appl
33	31	66.0	181	2	US-08-801-499-15	Sequence 15, Appl
34	31	66.0	181	2	US-08-615-271-15	Sequence 15, Appl
35	31	66.0	181	3	US-09-074-660-15	Sequence 15, Appl
36	31	66.0	181	3	US-09-074-659-15	Sequence 15, Appl
37	31	66.0	181	4	US-09-106-468-15	Sequence 15, Appl
38	31	66.0	181	4	US-09-106-466A-15	Sequence 15, Appl
39	31	66.0	181	4	US-09-106-467-15	Sequence 15, Appl
40	31	66.0	810	2	US-08-820-170A-14	Sequence 34, Appl
41	31	66.0	810	3	US-09-035-659-34	Sequence 34, Appl
42	31	66.0	810	4	US-09-273-565-34	Sequence 34, Appl
43	31	66.0	834	2	US-08-677-734A-9	Sequence 9, Appl
44	31	66.0	834	2	US-08-677-734A-10	Sequence 10, Appl
45	31	66.0	2020	1	US-07-551-531-2	Sequence 2, Appl

## ALIGNMENTS

RESULT 1

US-08-975-114A-4

Sequence 4, Application US/08975114A

Patent No. 5876714

GENERAL INFORMATION:

APPLICANT: Atsushi NISHIKAWA et al.

TITLE OF INVENTION: HUMAN GLYCOSYLTRANSFERASE GENE,

TITLE OF INVENTION: COMPOUNDS AND METHOD FOR INHIBITING CANCEROUS METASTASIS

NUMBER OF SEQUENCES: 5

CORRESPONDENCE ADDRESSES:

ADDRESSEE: Menderoth, Lind & Ponack, L.L.P.

STREET: 2033 K Street, N.W., Suite 800

CITY: Washington

STATE: D.C.

COUNTRY: U.S.A.

ZIP: 20006

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.5 Inch, 1.44 mb

COMPUTER: IBM Compatible

OPERATING SYSTEM: MS-DOS

SOFTWARE: Wordperfect 5.1

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/975,114A

FILING DATE: No. 5876714ember 20, 1997

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/107,173

FILING DATE: August 17, 1993

ATTORNEY/AGENT INFORMATION:

NAME: Lee Cheng

REGISTRATION NUMBER: 40,949

REFERENCE/DOCKET NUMBER: 1-F3439D1V

TELECOMMUNICATION INFORMATION:

TELEPHONE: 202-721-8200

TELEFAX: 202-721-8250

TELEX:

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 531 amino acid residues

TYPE: amino acid

STRANDEDNESS:

TOPOLOGY: linear

MOLECULE TYPE:

US-08-975-114A-4

Query Match 80.9%; Score 38; DB 2; Length 531;  
Best Local Similarity 75.0%; Pred. No. 66;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 GGVCVOPG 8  
DB 112 GGVCFKPG 119

## RESULT 2

US-08-849-281A-4  
Sequence 4, Application US/08849281A  
Patent No. 6153433  
GENERAL INFORMATION:  
APPLICANT: Eiji MIYOSHI et al.  
TITLE OF INVENTION: INHIBITOR FOR VIRAL REPLICATION  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Wenderoth, Lind & Ponack, L.L.P.  
STREET: 2033 K Street, N.W., Suite 800  
CITY: Washington  
STATE: D.C.  
COUNTRY: U.S.A.  
ZIP: 20006  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: Wordperfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/849,281A  
FILING DATE: May 30, 1997  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Lee Cheng  
REGISTRATION NUMBER: 40,949  
REFERENCE/DOCKET NUMBER: 97-0529\*/LC(MJ)/1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-721-8200  
TELEFAX: 202-721-8250  
MOLECULE TYPE: linear  
SEQUENCE CHARACTERISTICS:  
LENGTH: 531 amino acid residues  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: polypeptide  
US-08-849-281A-4

Query Match 80.9%; Score 38; DB 4; Length 531;  
Best Local Similarity 75.0%; Pred. No. 66;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 GGVCVOPG 8  
DB 112 GGVCFKPG 119

## RESULT 3

US-08-975-114A-5  
Sequence 5, Application US/08975114A  
Patent No. 5876714  
GENERAL INFORMATION:  
APPLICANT: Atsushi NISHIKAWA et al.  
TITLE OF INVENTION: HUMAN GLYCOSYLTRANSFERASE GENE,  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Wenderoth, Lind & Ponack, L.L.P.  
STREET: 2033 K Street, N.W., Suite 800  
CITY: Washington  
STATE: D.C.

COUNTRY: U.S.A.  
ZIP: 20006  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: Wordperfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/975,114A  
FILING DATE: No. 5876714member 20, 1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/107,173  
FILING DATE: August 17, 1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Lee Cheng  
REGISTRATION NUMBER: 40,949  
REFERENCE/DOCKET NUMBER: 1-F3439D1V  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-721-8200  
TELEFAX: 202-721-8250  
MOLECULE TYPE: linear  
SEQUENCE CHARACTERISTICS:  
LENGTH: 536 amino acid residues  
TYPE: amino acid  
STRANDEDNESS: linear  
TOPOLOGY: linear  
MOLECULE TYPE: linear  
US-08-975-114A-5

Query Match 80.9%; Score 38; DB 2; Length 536;  
Best Local Similarity 75.0%; Pred. No. 67;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 GGVCVOPG 8  
DB 112 GGVCFKPG 119

RESULT 4  
US-08-849-281A-3  
Sequence 3, Application US/08849281A  
Patent No. 6153433  
GENERAL INFORMATION:  
APPLICANT: Eiji MIYOSHI et al.  
TITLE OF INVENTION: INHIBITOR FOR VIRAL REPLICATION  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Wenderoth, Lind & Ponack, L.L.P.  
STREET: 2033 K Street, N.W., Suite 800  
CITY: Washington  
STATE: D.C.  
COUNTRY: U.S.A.  
ZIP: 20006  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: Wordperfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/849,281A  
FILING DATE: May 30, 1997  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Lee Cheng  
REGISTRATION NUMBER: 40,949  
REFERENCE/DOCKET NUMBER: 97-0529\*/LC(MJ)/1  
TELECOMMUNICATION INFORMATION:

TELEPHONE: 202-721-8200  
TELEFAX: 202-721-8250  
TELEX:  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 536 amino acid residues  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: polypeptide  
US-08-849-281A-3

Query Match 80.9%; Score 38; DB 4; Length 536;  
Best Local Similarity 75.0%; Pred. No. 67;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGVGVOPG 8  
DB 112 GGVGVOPG 119

RESULT 5  
US-08-859-931A-4  
Sequence 4, Application US/08859931A  
Patent No. 5945510  
GENERAL INFORMATION:  
APPLICANT: FASANO, Alessio  
TITLE OF INVENTION: SUBSTANTIALLY PURE ZONULIN, A  
TITLE OF INVENTION: PHYSIOLOGICAL MODULATOR OF  
TITLE OF INVENTION: MAMMALIAN TIGHT JUNCTIONS  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SUGHRUE, MION, ZINN, MACPEAK & SEAS  
STREET: 2100 Pennsylvania Avenue, N.W., Suite 800  
CITY: Washington, D.C.  
STATE: D.C.  
COUNTRY: U.S.A.  
ZIP: 20037  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/859,931A  
FILING DATE: 21 MAY 1997  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: KIT, Gordon  
REGISTRATION NUMBER: 30,764  
REFERENCE/DOCKET NUMBER: A-6901  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 293-7060  
TELEFAX: (202) 293-7860  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 20 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
HYPOTHETICAL: NO  
US-08-859-931A-4

Query Match 78.7%; Score 37; DB 2; Length 20;  
Best Local Similarity 87.5%; Pred. No. 4; 4;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGVGVOPG 8  
DB 8 GGVGVOPG 15

RESULT 6  
US-08-545-809A-125  
Sequence 125, Application US/08545809A  
Patent No. 6096878  
GENERAL INFORMATION:  
APPLICANT: Honjo, Tasuku  
TITLE OF INVENTION: HUMAN IMMUNOGLOBULIN VH GENE  
TITLE OF INVENTION: SEGMENTS AND DNA FRAGMENTS CONTAINING THE SAME  
NUMBER OF SEQUENCES: 145  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson, P.C.  
STREET: 225 Franklin Street  
CITY: Boston  
STATE: MA  
COUNTRY: US  
ZIP: 02110-2804  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM compatible  
OPERATING SYSTEM: Windows95  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/545,809A  
FILING DATE: 27-MAR-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/JP93/00603  
FILING DATE: 10-MAY-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Freeman, John W.  
REGISTRATION NUMBER: 29,066  
REFERENCE/DOCKET NUMBER: 06501/004001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-542-5070  
TELEFAX: 617-542-8906  
TELEX: 200154  
INFORMATION FOR SEQ ID NO: 125:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 118 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-545-809A-125

Query Match 78.7%; Score 37; DB 3; Length 118;  
Best Local Similarity 87.5%; Pred. No. 23;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGVGVOPG 8  
DB 27 GGVGVOPG 34

RESULT 7  
US-07-728-215-39  
Sequence 39, Application US/07728215  
Patent No. 5962643  
GENERAL INFORMATION:  
APPLICANT: Sheppard, Dean  
APPLICANT: Quaranta, Vito  
APPLICANT: Pytela, Robert  
TITLE OF INVENTION: A No. 5962643el Integrin Beta Subunit and Uses  
TITLE OF INVENTION: Thereof  
NUMBER OF SEQUENCES: 43  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pretly, Schroeder, Brueggemann & Clark  
STREET: 4370 La Jolla Village Drive, Suite 700  
CITY: San Diego  
STATE: California  
COUNTRY: United States of America  
ZIP: 92122

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/728,215  
FILING DATE: 19910711  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Campbell, Cathryn A.  
REGISTRATION NUMBER: 31,815  
REFERENCE/DOCKET NUMBER: P31 8717  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 535-9001  
TELEFAX: (619) 535-8949  
INFORMATION FOR SEQ ID NO: 39:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 92 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-07-728-215-39

Query Match 72.3%; Score 34; DB 2; Length 92;  
Best Local Similarity 83.3%; Pred. No. 54;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 3 VCVOPG 8  
Db 82 VCIOPG 87

RESULT 8  
US-07-728-215-41  
Sequence 41, Application US/07728215  
Patent No. 5962643  
GENERAL INFORMATION:  
APPLICANT: Sheppard, Dean  
APPLICANT: Quaranta, Wilo  
APPLICANT: Pytela, Robert  
TITLE OF INVENTION: A No. 5962643el Integrin Beta Subunit and Uses  
TITLE OF INVENTION: Tereof  
NUMBER OF SEQUENCES: 13  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pretty, Schroeder, Brueggemann & Clark  
STREET: 4370 La Jolla Village Drive, Suite 700  
CITY: San Diego  
STATE: California  
COUNTRY: United States of America  
ZIP: 92122  
COMPUTER READABLE FORM:  
MEDIUM TYPE: IBM PC compatible  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/728,215  
FILING DATE: 19910711  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Campbell, Cathryn A.  
REGISTRATION NUMBER: 31,815  
REFERENCE/DOCKET NUMBER: P31 8717  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 535-9001  
TELEFAX: (619) 535-8949  
INFORMATION FOR SEQ ID NO: 41:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 92 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: linear

MOLECULE TYPE: protein  
US-07-728-215-41

Query Match 72.3%; Score 34; DB 2; Length 92;  
Best Local Similarity 83.3%; Pred. No. 54;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 3 VCVOPG 8  
Db 82 VCIOPG 87

RESULT 9  
US-08-444-792-4  
Sequence 4, Application US/08444792  
Patent No. 5726037  
GENERAL INFORMATION:  
APPLICANT: Bodary, Sarah C.  
APPLICANT: Gorman, Cornelia M.  
APPLICANT: McLean, John W.  
APPLICANT: Napier, Mary A.  
TITLE OF INVENTION: METHOD FOR PREPARING WATER SOLUBLE  
NUMBER OF SEQUENCES: 18  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genentech, Inc.  
STREET: 460 Point San Bruno Blvd  
CITY: South San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94080  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 Inch, 720 Kb floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: WinPatIn (Genentech)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/444,792  
FILING DATE: 19-May-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/380227  
FILING DATE: 30-JAN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/218878  
FILING DATE: 28-MAR-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/821337  
FILING DATE: 13-JAN-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/444490  
FILING DATE: 01-DEC-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/290224  
FILING DATE: 22-DEC-1988  
ATTORNEY/AGENT INFORMATION:  
NAME: Lee, Wendy M.  
REGISTRATION NUMBER: 00,000  
REFERENCE/DOCKET NUMBER: P055221C3D4  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415/225-1994  
TELEFAX: 415/952-9881  
TELEX: 910/371-7168  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 718 amino acids  
TYPE: Amino Acid  
TOPOLOGY: Linear  
US-08-444-792-4

Query Match 72.3%; Score 34; DB 1; Length 718;



Best Local Similarity 83.3%, Pred. No. 3.7e+02;  
Matches 5: Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 VCVOPG 8  
11:1111  
Db 613 VCIOPG 618

RESULT 10  
US-08-445-042-4  
: Sequence 4, Application US/08445042  
: Patent No. 5726290  
: GENERAL INFORMATION:  
: APPLICANT: Bodary, Sarah C.  
: APPLICANT: Gorman, Cornelia M.  
: APPLICANT: McLean, John W.  
: APPLICANT: Napier, Mary A.  
: TITLE OF INVENTION: METHOD FOR PREPARING WATER SOLUBLE  
: POLYPEPTIDES  
: NUMBER OF SEQUENCES: 18  
: CORRESPONDENCE ADDRESS:  
: ADDRESSEE: Genentech, Inc.  
: STREET: 460 Point San Bruno Blvd  
: CITY: South San Francisco  
: STATE: California  
: COUNTRY: USA  
: ZIP: 94080  
: COMPUTER READABLE FORM:  
: MEDIUM TYPE: 3.5 inch, 720 Kb floppy disk  
: COMPUTER: IBM PC compatible  
: OPERATING SYSTEM: PC-DOS/MS-DOS  
: SOFTWARE: Winpatln (Genentech)  
: CURRENT APPLICATION DATA:  
: APPLICATION NUMBER: US/08/445.042  
: FILING DATE: 19-May-1995  
: CLASSIFICATION: 530  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: 08/380227  
: FILING DATE: 30-JAN-1995  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: 08/218678  
: FILING DATE: 28-MAR-1994  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: 07/821337  
: FILING DATE: 13-JAN-1992  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: 07/444490  
: FILING DATE: 01-DEC-1989  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: 07/290224  
: FILING DATE: 22-DEC-1988  
: ATTORNEY/AGENT INFORMATION:  
: NAME: Lee, Wendy M.  
: REGISTRATION NUMBER: 00,000  
: REFERENCE/DOCKET NUMBER: P0552P1C3D1  
: TELECOMMUNICATION INFORMATION:  
: TELEPHONE: 415/225-1994  
: TELEFAX: 415/952-9881  
: TELEFAX: 910/371-7168  
: INFORMATION FOR SEQ ID NO: 4:  
: SEQUENCE CHARACTERISTICS:  
: LENGTH: 718 amino acids  
: TYPE: Amino Acid  
: TOPOLOGY: Linear  
: US-08-445-042-4

Query Match 72.3%, Score 34; DB 1; Length 718;  
Best Local Similarity 83.3%, Pred. No. 3.7e+02;  
Matches 5: Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 3 VCVOPG 8  
11:1111

Db 613 VCIOPG 618

RESULT 11  
US-07-728-215-32  
: Sequence 32, Application US/07728215  
: Patent No. 5962643  
: GENERAL INFORMATION:  
: APPLICANT: Sheppard, Dean  
: APPLICANT: Quaranta, Vito  
: APPLICANT: Pytela, Robert  
: TITLE OF INVENTION: A No. 5962643el Integrin Beta Subunit and Uses  
: THEREOF  
: NUMBER OF SEQUENCES: 43  
: CORRESPONDENCE ADDRESS:  
: ADDRESSEE: Pretly, Schroeder, Brueggemann & Clark  
: STREET: 4370 La Jolla Village Drive, Suite 700  
: CITY: San Diego  
: STATE: California  
: COUNTRY: United States of America  
: ZIP: 92122  
: COMPUTER READABLE FORM:  
: MEDIUM TYPE: Floppy disk  
: COMPUTER: IBM PC compatible  
: OPERATING SYSTEM: PC-DOS/MS-DOS  
: SOFTWARE: Patentln Release #1.0, Version #1.25  
: CURRENT APPLICATION DATA:  
: APPLICATION NUMBER: US/07/728.215  
: FILING DATE: 19910711  
: CLASSIFICATION: 435  
: ATTORNEY/AGENT INFORMATION:  
: NAME: Campbell, Cathryn A.  
: REGISTRATION NUMBER: 31,815  
: REFERENCE/DOCKET NUMBER: P31 8717  
: TELECOMMUNICATION INFORMATION:  
: TELEPHONE: (619) 535-9001  
: TELEFAX: (619) 535-8949  
: INFORMATION FOR SEQ ID NO: 32:  
: SEQUENCE CHARACTERISTICS:  
: LENGTH: 788 amino acids  
: TYPE: AMINO ACID  
: STRANDEDNESS: single  
: TOPOLOGY: linear  
: US-07-728-215-32

Query Match 72.3%, Score 34; DB 2; Length 788;  
Best Local Similarity 83.3%, Pred. No. 4e+02;  
Matches 5: Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 VCVOPG 8  
11:1111  
Db 613 VCIOPG 618

RESULT 12  
US-08-545-809A-108  
: Sequence 108, Application US/08545809A  
: Patent No. 6096878  
: GENERAL INFORMATION:  
: APPLICANT: Honjo, Tasuku  
: APPLICANT: Matsuda, Fumihiko  
: TITLE OF INVENTION: HUMAN IMMUNOGLOBULIN VH GENE  
: SEGMENTS AND DNA FRAGMENTS CONTAINING THE SAME  
: NUMBER OF SEQUENCES: 145  
: CORRESPONDENCE ADDRESS:  
: ADDRESSEE: Fish & Richardson, P.C.  
: STREET: 225 Franklin Street  
: CITY: Boston  
: STATE: MA  
: COUNTRY: US  
: ZIP: 02110-2804  
: COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: Windows95  
SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/545,809A  
FILING DATE: 27-MAR-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/JP93/00603  
FILING DATE: 10-MAY-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Freeman, John W.  
REGISTRATION NUMBER: 29,066  
REFERENCE/DOCKET NUMBER: 06501/004001  
TELEPHONE: 617-542-5070  
TELEFAX: 617-542-8906  
TELEX: 200154  
INFORMATION FOR SEQ ID NO: 108:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 118 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-545-809A-108

Query Match 70.2%; Score 33; DB 3; Length 118;  
Best Local Similarity 75.0%; Pred. No. 98;  
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 GGCVCOPG 8  
Db 27 GGALVPRG 34

RESULT 13  
US-08-475-000-16  
Sequence 16, Application US/08475000  
Patent No. 5811267  
GENERAL INFORMATION:  
APPLICANT: RING, DAVID B.  
TITLE OF INVENTION: ANTIGEN-BINDING SITES OF ANTIBODY  
TITLE OF INVENTION: MOLECULES SPECIFIC FOR CANCER ANTIGENS  
NUMBER OF SEQUENCES: 18  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CHIRON CORPORATION  
STREET: INTELLECTUAL PROPERTY R440 - P.O. BOX 8097  
CITY: Emeryville  
STATE: CA  
COUNTRY: USA  
ZIP: 94662-8097  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/475,000  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: SAVERIDE, PAUL  
REGISTRATION NUMBER: 36,914  
REFERENCE/DOCKET NUMBER: 0850,007  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (510) 601-2585  
TELEFAX: (510) 655-3542  
INFORMATION FOR SEQ ID NO: 16:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 119 amino acids  
TYPE: amino acid  
TOPOLOGY: linear

MOLECULE TYPE: protein  
US-08-475-000-16

Query Match 70.2%; Score 33; DB 2; Length 119;  
Best Local Similarity 75.0%; Pred. No. 98;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 GGCVCOPG 8  
Db 8 GGALVPRG 15

RESULT 14  
US-08-483-199-16  
Sequence 16, Application US/08483199  
Patent No. 5849877  
GENERAL INFORMATION:  
APPLICANT: RING, DAVID B.  
TITLE OF INVENTION: ANTIGEN-BINDING SITES OF ANTIBODY  
TITLE OF INVENTION: MOLECULES SPECIFIC FOR CANCER ANTIGENS  
NUMBER OF SEQUENCES: 18  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CHIRON CORPORATION  
STREET: INTELLECTUAL PROPERTY - R440, P.O. BOX 8097  
CITY: Emeryville  
STATE: CA  
COUNTRY: USA  
ZIP: 94662-8097  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/483,199  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: SAVERIDE, PAUL  
REGISTRATION NUMBER: 36,914  
REFERENCE/DOCKET NUMBER: 0850,009  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (510) 601-2585  
TELEFAX: (510) 655-3542  
INFORMATION FOR SEQ ID NO: 16:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 119 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-483-199-16

Query Match 70.2%; Score 33; DB 2; Length 119;  
Best Local Similarity 75.0%; Pred. No. 98;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 GGCVCOPG 8  
Db 8 GGALVPRG 15

RESULT 15  
US-08-484-508-16  
Sequence 16, Application US/08484508  
Patent No. 5948647  
GENERAL INFORMATION:  
APPLICANT: RING, DAVID B.  
TITLE OF INVENTION: ANTIGEN-BINDING SITES OF ANTIBODY  
TITLE OF INVENTION: MOLECULES SPECIFIC FOR CANCER ANTIGENS  
NUMBER OF SEQUENCES: 18  
CORRESPONDENCE ADDRESS:

ADDRESSEE: CHIRON CORPORATION  
STREET: INTELLECTUAL PROPERTY - R440, P.O. BOX 8097  
CITY: Emeryville  
STATE: CA  
COUNTRY: USA  
ZIP: 94662-8097  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/484,508  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:  
NAME: SAVEREIDE, PAUL  
REGISTRATION NUMBER: 36,914  
REFERENCE/DOCKET NUMBER: 0850.008  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (510) 601-585  
TELEFAX: (510) 655-3542  
INFORMATION FOR SEQ ID NO: 16:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 119 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-484-508-16

Query Match 70.2%; Score 33; DB 2; Length 119;  
Best Local Similarity 75.0%; Pred. No. 98;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
QY 1 GGYCVCPG 8  
||| ||||  
Db 8 GGYLVPRG 15

Search completed: June 13, 2001, 14:16:35  
Job time: 496 sec

*This Page Blank (uspto)*

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

## OM protein - protein search, using sw model

Run on: June 13, 2001, 14:14:36 ; Search time 229.28 Seconds

(without alignments)  
1.995 Million cell updates/sec

Title: PCT-US01-05825A-14

Perfect score: 46  
Sequence: 1 GCVCVQDC 8

## Scoring table:

BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 390729 seqs, 57163235 residues

Total number of hits satisfying chosen parameters: 390729

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 08  
Maximum Match 1008

Listing first 45 summaries

## Database :

A.GeneSeq\_0401.\*  
1: /SID56/gcgdata/geneseq/geneseq/AA1980.DAT.\*  
2: /SID56/gcgdata/geneseq/geneseq/AA1981.DAT.\*  
3: /SID56/gcgdata/geneseq/geneseq/AA1982.DAT.\*  
4: /SID56/gcgdata/geneseq/geneseq/AA1983.DAT.\*  
5: /SID56/gcgdata/geneseq/geneseq/AA1984.DAT.\*  
6: /SID56/gcgdata/geneseq/geneseq/AA1985.DAT.\*  
7: /SID56/gcgdata/geneseq/geneseq/AA1986.DAT.\*  
8: /SID56/gcgdata/geneseq/geneseq/AA1987.DAT.\*  
9: /SID56/gcgdata/geneseq/geneseq/AA1988.DAT.\*  
10: /SID56/gcgdata/geneseq/geneseq/AA1989.DAT.\*  
11: /SID56/gcgdata/geneseq/geneseq/AA1990.DAT.\*  
12: /SID56/gcgdata/geneseq/geneseq/AA1991.DAT.\*  
13: /SID56/gcgdata/geneseq/geneseq/AA1992.DAT.\*  
14: /SID56/gcgdata/geneseq/geneseq/AA1993.DAT.\*  
15: /SID56/gcgdata/geneseq/geneseq/AA1994.DAT.\*  
16: /SID56/gcgdata/geneseq/geneseq/AA1995.DAT.\*  
17: /SID56/gcgdata/geneseq/geneseq/AA1996.DAT.\*  
18: /SID56/gcgdata/geneseq/geneseq/AA1997.DAT.\*  
19: /SID56/gcgdata/geneseq/geneseq/AA1998.DAT.\*  
20: /SID56/gcgdata/geneseq/geneseq/AA1999.DAT.\*  
21: /SID56/gcgdata/geneseq/geneseq/AA2000.DAT.\*  
22: /SID56/gcgdata/geneseq/geneseq/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	46	100.0	8	21	Y79118	Peptide antagonist
2	43	93.5	8	21	Y79122	Peptide antagonist
3	41	89.1	196	20	Y36831	Amino acid sequenc
4	39	84.8	8	21	Y79117	Peptide antagonist
5	39	84.8	8	21	Y79126	Peptide antagonist
6	38	82.6	8	21	Y79106	Peptide antagonist
7	37	80.4	380	18	W09406	Transforming growt
8	36	78.3	8	21	Y79120	Peptide antagonist
9	36	78.3	8	21	Y79121	Peptide antagonist
10	36	78.3	115	20	Y42459	Human guanylin pre
11	36	78.3	115	20	Y42460	Human guanylin pre

12	36	78.3	115	20	Y27892	Human secreted pro
13	36	78.3	336	17	W05520	HCW Toledo strain
14	35	76.1	8	21	Y79110	Peptide antagonist
15	35	76.1	399	13	R20006	Zonula occludens t
16	34	73.9	449	20	Y34458	Porphyromonas ging
17	34	73.9	467	20	Y34334	Porphyromonas ging
18	34	73.9	473	17	R86869	Adhesive protein.
19	34	73.9	563	18	W34674	Human mammary carc
20	34	73.9	624	21	B34286	Human pancreatic c
21	34	73.9	669	19	W70893	Maize starch solub
22	34	73.9	669	19	W56486	Zea mays soluble s
23	34	73.9	792	16	R85198	Avenacinase - a sa
24	34	73.9	793	16	R85200	Avenacinase-like p
25	34	73.9	793	16	R85199	Human mammary carc
26	34	73.9	876	18	W34675	Human mammary carc
27	34	73.9	882	18	W34673	Human mammary carc
28	34	73.9	914	16	R71100	Protein-tyrosine-k
29	34	73.9	919	16	R75502	Human mammary carc
30	34	73.9	919	16	R75504	Human mammary carc
31	34	73.9	919	18	W34672	Human mammary carc
32	34	73.9	1487	19	W61562	Human type II coll
33	33	71.7	8	21	Y79124	Peptide antagonist
34	33	71.7	31	20	W88384	Human Zneu1 EGF-11
35	33	71.7	73	20	W88389	Human Zneu1 partia
36	33	71.7	115	21	B41718	Human OREF ORF1402
37	33	71.7	153	21	B41638	Human Zneu1 partia
38	33	71.7	158	20	W88388	Human Zneu1 partia
39	33	71.7	169	20	W88390	Human Zneu1 partia
40	33	71.7	224	20	Y59870	Human normal uteru
41	33	71.7	247	21	Y52139	Human TANGO 125b (
42	33	71.7	254	20	W88382	Human neuro-growth
43	33	71.7	265	21	B42204	Human OREF ORF1968
44	33	71.7	273	20	Y41769	Human PRO213-1 pro
45	33	71.7	273	20	Y41770	Human PRO1330 prot

## ALIGNMENTS

RESULT	1
ID	Y79118
Y79118	standard; Peptide; 8 AA.
XX	XX
AC	Y79118;
DF	05-JUN-2000 (first entry)
XX	XX
DE	Peptide antagonist of zonulin.
XX	XX
KW	Zonulin; antagonist; zonula occludens toxin receptor;
KW	blood-brain barrier; antiinflammatory; cerebroprotective;
KW	neuroprotective; dermatological; antidiabetic; antiviral;
KW	antibacterial; cytostatic; anti-HIV; vulnery; antiallergic;
KW	hypotensive; immunosuppressive; antiparasitic; vasotropic;
KW	gastrointestinal inflammation; therapy.
XX	XX
OS	Synthetic.
XX	XX
PN	W0200007609-A1.
XX	XX
PD	17-FEB-2000.
XX	XX
PF	28-JUL-1999; 99WO-US16683.
XX	XX
PR	03-AUG-1998; 98US-0127815.
XX	XX
PA	(UYMA-) UNIV MARYLAND BALTIMORE.
XX	XX
PI	Fasano A;
XX	XX
DR	WPI; 2000-205565/18.
XX	XX
PT	New peptide antagonist of zonulin useful as antiinflammatory agent for

PT treating cerebral ischemia, stroke, cerebral edema, gastritis,  
 PT shigellosis, viral gastroenteritis, meningitis, encephalomyelitis  
 PS  
 XX  
 Claim 1; Page 45; 69pp; English.

CC This present sequence is that of a peptide antagonist of zonulin  
 CC (2), one of 25 such peptides (see Y79105-29) of the invention,  
 CC which bind to a zonula occludens toxin (ZOT) receptor, yet do not  
 CC physiologically modulate the opening of mammalian tight junctions  
 CC (TJ). The peptide antagonists are based on a common motif of ZOT  
 CC and human zonulins, which is believed to be critical for receptor  
 CC binding. They can be prepared by chemical synthesis or by use of  
 CC recombinant DNA techniques. The peptide antagonists are used as an  
 CC antiinflammatory agents in the treatment of gastrointestinal  
 CC inflammation, where they bind to the ZOT receptor in the intestine  
 CC and yet does not physiologically modulate the opening of TJ in the  
 CC intestine. Gastrointestinal inflammation conditions give rise to  
 CC increased intestinal permeability and the peptide is useful for  
 CC treating intestinal conditions that cause protein losing enteropathy  
 CC caused by infection, e.g. Clostridium difficile infection,  
 CC enterocolitis, shigellosis, viral gastroenteritis, parasite  
 CC infestation, bacterial overgrowth, whipple's disease, diseases with  
 CC mucosal erosion or ulcerations, e.g. gastritis, gastric cancer,  
 CC collagenous colitis, inflammatory bowel disease, diseases marked by  
 CC lymphatic obstruction, e.g. congenital intestinal lymphangiectasia,  
 CC sarcoidosis lymphoma, mesenteric tuberculosis, and after surgical  
 CC correction of congenital heart disease with Fontan's operation,  
 CC mucosal diseases without ulceration, e.g. Menetrier's disease,  
 CC coeliac disease, eosinophilic gastroenteritis, and immune diseases,  
 CC e.g. systemic lupus erythematosus or food allergies, primarily to  
 CC milk.

SO Sequence 8 AA:

Query Match 100.0%; Score 46; DB 21; Length 8;  
 Best Local Similarity 100.0%; Pred. No. 3.2e+05;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GGYCVQDG 8

DB 1 ggycvqdg 8

RESULT 2

Y79122  
 ID Y79122 standard; Peptide: 8 AA.

XX Y79122;

DE 05-JUN-2000 (first entry)

XX Peptide antagonist of zonulin.

KW Zonulin; antagonist; zonula occludens toxin receptor;  
 KW blood-brain barrier; antiinflammatory; cerebroprotective;  
 KW neuroprotective; dermatological; antilucer; antiviral;  
 KW antibacterial; cytostatic; anti-HIV; vulnery; antiallergic;  
 KW hypotensive; immunosuppressive; antiparasitic; vasotropic;  
 KW gastrointestinal inflammation; therapy.

OS Synthetic.

PN WO200007609-A1.

PD 17-FEB-2000.

PF 28-JUL-1999; 99WO-US16683.

PR 03-AUG-1998; 98US-0127815.

PA (UYMA-) UNIV MARYLAND BALTIMORE.

PI Fasano A;  
 XX WPI: 2000-205565/18.  
 DR  
 XX

PT New peptide antagonist of zonulin useful as antiinflammatory agent for  
 PT treating cerebral ischemia, stroke, cerebral edema, gastritis,  
 PT shigellosis, viral gastroenteritis, meningitis, encephalomyelitis  
 PS  
 XX  
 Claim 1; Page 46; 69pp; English.

CC This present sequence is that of a peptide antagonist of zonulin  
 CC (2), one of 25 such peptides (see Y79105-29) of the invention,  
 CC which bind to a zonula occludens toxin (ZOT) receptor, yet do not  
 CC physiologically modulate the opening of mammalian tight junctions  
 CC (TJ). The peptide antagonists are based on a common motif of ZOT  
 CC and human zonulins, which is believed to be critical for receptor  
 CC binding. They can be prepared by chemical synthesis or by use of  
 CC recombinant DNA techniques. The peptide antagonists are used as an  
 CC antiinflammatory agents in the treatment of gastrointestinal  
 CC inflammation, where they bind to the ZOT receptor in the intestine  
 CC and yet does not physiologically modulate the opening of TJ in the  
 CC intestine. Gastrointestinal inflammation conditions give rise to  
 CC increased intestinal permeability and the peptide is useful for  
 CC treating intestinal conditions that cause protein losing enteropathy  
 CC caused by infection, e.g. Clostridium difficile infection,  
 CC enterocolitis, shigellosis, viral gastroenteritis, parasite  
 CC infestation, bacterial overgrowth, whipple's disease, diseases with  
 CC mucosal erosion or ulcerations, e.g. gastritis, gastric cancer,  
 CC collagenous colitis, inflammatory bowel disease, diseases marked by  
 CC lymphatic obstruction, e.g. congenital intestinal lymphangiectasia,  
 CC sarcoidosis lymphoma, mesenteric tuberculosis, and after surgical  
 CC correction of congenital heart disease with Fontan's operation,  
 CC mucosal diseases without ulceration, e.g. Menetrier's disease,  
 CC coeliac disease, eosinophilic gastroenteritis, and immune diseases,  
 CC e.g. systemic lupus erythematosus or food allergies, primarily to  
 CC milk.

SO Sequence 8 AA:

Query Match 93.5%; Score 43; DB 21; Length 8;  
 Best Local Similarity 87.5%; Pred. No. 3.2e+05;  
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GGYCVQDG 8

DB 1 ggycvqdg 8

RESULT 3

Y36831  
 ID Y36831 standard; Protein; 196 AA.

XX Y36831;

DE 07-OCT-1999 (first entry)

XX Amino acid sequence of a Chlamydia trachomatis protein.

KW Vaccine; eye disease; conventional trachoma; nonendemic trachoma;  
 KW paratrachoma; inclusion conjunctivitis; genital disease; perihepatitis;  
 KW nongonococcal urethritis; epididymitis; cervicitis; salpingitis;  
 KW bartolinitis; pneumopathy; venereal lymphogranulomatosis.

OS Chlamydia trachomatis.

PN WO9928475-A2.

PD 10-JUN-1999.

PF 27-NOV-1998; 98WO-IB01939.

PR 04-NOV-1998; 98US-0107077.

PR 28-NOV-1997: 97FR-0015041.  
 PR 17-DEC-1997: 97FR-0016034.  
 XX (GEST ) GENSET.  
 PA  
 XX  
 PI Griffiths R;  
 XX  
 DR WPI: 1999-371125/31.  
 XX  
 PT Genome sequence of Chlamydia trachomatis  
 XX  
 PS Disclosure: Page 710-711: 1755pp: English.  
 XX  
 CC Y36754-Y37949 are encoded by open reading frames (ORFs) of the genome of  
 CC Chlamydia trachomatis (see Z01425). The polypeptides can be used as  
 CC vaccines against Chlamydia trachomatis. Antisense and ribozyme sequences  
 CC can also be used to control growth of the microorganism. Chlamydia  
 CC trachomatis is responsible for a large number of diseases, e.g. eye  
 CC diseases such as conventional trachoma, nongonemic trachoma,  
 CC paratrachoma, and inclusion conjunctivitis; genital diseases such as  
 CC nongonococcal urethritis, epididymitis, cervicitis, salpingitis,  
 CC peritonitis, Bartholinitis; pneumonia in breast feeding infants;  
 CC and venereal lymphogranulomatosis. The polypeptides of the invention  
 CC may be of use in treating these diseases.  
 CC  
 SO Sequence 196 AA;  
 Query Match 89.1%; Score 41; DB 20; Length 196;  
 Best Local Similarity 87.5%; Pred. No. 13;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 OY 1 GGCVCQDC 8  
 DB 150 ggcvcvqdg 157  
 RESULT 4  
 Y79117  
 ID Y79117 standard; Peptide; 8 AA.  
 AC Y79117;  
 XX  
 DT 05-JUN-2000 (first entry)  
 XX  
 DE Peptide antagonist of zonulin.  
 XX  
 KW Zonulin; antagonist; zonula occludens toxin receptor;  
 KW blood-brain barrier; antiinflammatory; cerebroprotective;  
 KW neuroprotective; dermatological; antiulcer; antiviral;  
 KW antibacterial; cytostatic; anti-HIV; vulnery; antiallergic;  
 KW hypotensive; immunosuppressive; antiparasitic; vasotropic;  
 KW gastrointestinal inflammation; therapy.  
 KW  
 XX  
 OS Synthetic.  
 XX  
 PN WO200007609-A1.  
 PD 17-FEB-2000.  
 XX  
 PF 28-JUL-1999; 99WO-US16683.  
 XX  
 PR 03-AUG-1998; 98US-0127815.  
 XX  
 PA (UYMA-) UNIV MARYLAND BALTIMORE.  
 XX  
 PI Fasano A;  
 XX  
 DR WPI: 2000-205565/18.  
 XX  
 PT New peptide antagonist of zonulin useful as antiinflammatory agent for  
 PT treating cerebral ischemia, stroke, cerebral edema, gastritis,  
 PT shigellosis, viral gastroenteritis, meningitis, encephalomyelitis -

XX  
 PS Claim 1: Page 45: 69pp: English.  
 XX  
 CC This present sequence is that of a peptide antagonist of zonulin  
 CC (2), one of 25 such peptides (see Y79105-29) of the invention,  
 CC which bind to a zonula occludens toxin (ZOT) receptor. Yet do not  
 CC physiologically modulate the opening of mammalian tight junctions  
 CC (TJ). The peptide antagonists are based on a common motif of ZOT  
 CC and human zonulins, which is believed to be critical for receptor  
 CC binding. They can be prepared by chemical synthesis or by use of  
 CC recombinant DNA techniques. The peptide antagonists are used as an  
 CC antiinflammatory agents in the treatment of gastrointestinal  
 CC inflammation, where they bind to the ZOT receptor in the intestine  
 CC and yet does not physiologically modulate the opening of TJ in the  
 CC intestine. Gastrointestinal inflammation conditions give rise to  
 CC increased intestinal permeability and the peptide is useful for  
 CC treating intestinal conditions that cause protein losing enteropathy  
 CC caused by infection, e.g. Clostridium difficile infection,  
 CC enterocolitis, shigellosis, viral gastroenteritis, parasite  
 CC infestation, bacterial overgrowth, whipple's disease, diseases with  
 CC mucosal erosion or ulcerations, e.g. gastritis, gastric cancer,  
 CC collagenous colitis, inflammatory bowel disease, diseases marked by  
 CC lymphatic obstruction, e.g. congenital intestinal lymphangiectasia,  
 CC sarcoidosis lymphoma, mesenteric tuberculosis, and after surgical  
 CC correction of congenital heart disease with Fontan's operation,  
 CC mucosal diseases without ulceration, e.g. Menetrier's disease,  
 CC coeliac disease, eosinophilic gastroenteritis, and immune diseases,  
 CC e.g. systemic lupus erythematosus or food allergies, primarily to  
 CC milk.  
 CC  
 SO Sequence 8 AA;  
 Query Match 84.8%; Score 39; DB 21; Length 8;  
 Best Local Similarity 87.5%; Pred. No. 3.2e+05;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 OY 1 GGCVCQDC 8  
 DB 1 ggcvcvqpg 8  
 RESULT 5  
 Y79126  
 ID Y79126 standard; Peptide; 8 AA.  
 AC Y79126;  
 XX  
 DT 05-JUN-2000 (first entry)  
 XX  
 DE Peptide antagonist of zonulin.  
 XX  
 KW Zonulin; antagonist; zonula occludens toxin receptor;  
 KW blood-brain barrier; antiinflammatory; cerebroprotective;  
 KW neuroprotective; dermatological; antiulcer; antiviral;  
 KW antibacterial; cytostatic; anti-HIV; vulnery; antiallergic;  
 KW hypotensive; immunosuppressive; antiparasitic; vasotropic;  
 KW gastrointestinal inflammation; therapy.  
 KW  
 XX  
 OS Synthetic.  
 XX  
 PN WO200007609-A1.  
 PD 17-FEB-2000.  
 XX  
 PF 28-JUL-1999; 99WO-US16683.  
 XX  
 PR 03-AUG-1998; 98US-0127815.  
 XX  
 PA (UYMA-) UNIV MARYLAND BALTIMORE.  
 XX  
 PI Fasano A;  
 XX

DR WPI; 2000-205565/18.

XX New peptide antagonist of zonulin useful as antiinflammatory agent for  
PT treating cerebral ischemia, stroke, cerebral edema, gastritis,  
PT shigellosis, viral gastroenteritis, meningitis, encephalomyelitis  
XX  
XX  
PS Claim 1; Page 48; 69pp; English.

CC This present sequence is that of a peptide antagonist of zonulin  
CC (2), one of 25 such peptides (see Y79105-29) of the invention,  
CC which bind to a zonula occludens toxin (ZOT) receptor, yet do not  
CC physiologically modulate the opening of mammalian tight junctions  
CC (TJ). The peptide antagonists are based on a common motif of ZOT  
CC and human zonulins, which is believed to be critical for receptor  
CC binding. They can be prepared by chemical synthesis or by use of  
CC recombinant DNA techniques. The peptide antagonists are used as an  
CC antiinflammatory agents in the treatment of gastrointestinal  
CC inflammation, where they bind to the ZOT receptor in the intestine  
CC and yet does not physiologically modulate the opening of TJ in the  
CC intestine. Gastrointestinal inflammation conditions give rise to  
CC increased intestinal permeability and the peptide is useful for  
CC treating intestinal conditions that cause protein losing enteropathy  
CC caused by infection, e.g. Clostridium difficile infection,  
CC enterocolitis, shigellosis, viral gastroenteritis, parasite  
CC infestation, bacterial overgrowth, whipple's disease, diseases with  
CC mucosal erosion or ulcerations, e.g. gastritis, gastric cancer,  
CC collagenous colitis, inflammatory bowel disease, diseases marked by  
CC lymphatic obstruction, e.g. congenital intestinal lymphangiectasia,  
CC sarcoïdosis lymphoma, mesenteric tuberculosis, and after surgical  
CC correction of congenital heart disease with Fontan's operation,  
CC mucosal diseases without ulceration, e.g. Menetrier's disease,  
CC coeliac disease, eosinophilic gastroenteritis, and immune diseases,  
CC e.g. systemic lupus erythematosus or food allergies, primarily to  
CC milk.

CC Sequence 8 AA:

Query Match Best Local Similarity 84.8%; Score 39; DB 21; Length 8;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGVGVQDG 8  
IIIIII

DB 1 ggvqvqdg 8  
Y79106 standard; Peptide; 8 AA.

DE Peptide antagonist of zonulin.

XX Zonulin; antagonist; zonula occludens toxin receptor;  
KW blood-brain barrier; antiinflammatory; cerebroprotective;  
KW neuroprotective; dermatological; antitumor; antiviral;  
KW antibacterial; cytostatic; anti-HIV; vulnereary; anti-allergic;  
KW hypotensive; immunosuppressive; antiparasitic; vasotropic;  
KW gastrointestinal inflammation; therapy.

XX Synthetic.

OS WO200007609-A1.

PN 17-FEB-2000.

PD 28-JUL-1999; 99WO-US16683.

PF 03-AUG-1998; 98US-0127815.

XX (UYMA-) UNIV MARYLAND BALTIMORE.

XX Pasano A;

XX WPI; 2000-205565/18.

XX New peptide antagonist of zonulin useful as antiinflammatory agent for  
PT treating cerebral ischemia, stroke, cerebral edema, gastritis,  
PT shigellosis, viral gastroenteritis, meningitis, encephalomyelitis  
XX  
XX  
PS Claim 1; Page 41; 69pp; English.

CC This present sequence is that of a peptide antagonist of zonulin  
CC (2), one of 25 such peptides (see Y79105-29) of the invention,  
CC which bind to a zonula occludens toxin (ZOT) receptor, yet do not  
CC physiologically modulate the opening of mammalian tight junctions  
CC (TJ). The peptide antagonists are based on a common motif of ZOT  
CC and human zonulins, which is believed to be critical for receptor  
CC binding. They can be prepared by chemical synthesis or by use of  
CC recombinant DNA techniques. The peptide antagonists are used as an  
CC antiinflammatory agents in the treatment of gastrointestinal  
CC inflammation, where they bind to the ZOT receptor in the intestine  
CC and yet does not physiologically modulate the opening of TJ in the  
CC intestine. Gastrointestinal inflammation conditions give rise to  
CC increased intestinal permeability and the peptide is useful for  
CC treating intestinal conditions that cause protein losing enteropathy  
CC caused by infection, e.g. Clostridium difficile infection,  
CC enterocolitis, shigellosis, viral gastroenteritis, parasite  
CC infestation, bacterial overgrowth, whipple's disease, diseases with  
CC mucosal erosion or ulcerations, e.g. gastritis, gastric cancer,  
CC collagenous colitis, inflammatory bowel disease, diseases marked by  
CC lymphatic obstruction, e.g. congenital intestinal lymphangiectasia,  
CC sarcoïdosis lymphoma, mesenteric tuberculosis, and after surgical  
CC correction of congenital heart disease with Fontan's operation,  
CC mucosal diseases without ulceration, e.g. Menetrier's disease,  
CC coeliac disease, eosinophilic gastroenteritis, and immune diseases,  
CC e.g. systemic lupus erythematosus or food allergies, primarily to  
CC milk.

CC Sequence 8 AA:

Query Match Best Local Similarity 82.6%; Score 38; DB 21; Length 8;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGVGVQDG 8  
IIIIII

DB 1 ggvqvqdg 8  
W09406 standard; Protein; 380 AA.

DE 03-SEP-1997 (first entry)

XX Transforming growth factor alpha-HI.

KW TGF-alpha; human; transforming growth factor alpha-HI; treat;

KW promote; wound healing; neurological functioning; trauma; AIDS dementia;

XX ocular disorder; susceptibility; diagnosis.

OS Homo sapiens.

XX Key

FT Peptide Location/Qualifiers  
FT 1..39  
FT /note="putative signal peptide"  
FT 40..266  
FT /note="putative precursor sequence"



FT Domain 267..316  
 FT /note= "putative soluble portion"  
 FT 317..380  
 FT /note= "putative transmembrane portion"  
 PN MO9639497-A1.  
 XX 12-DEC-1996.  
 PD 06-JUN-1996; 96WO-US09448.  
 PF 06-JUN-1996; 95US-0468846.  
 PR 06-JUN-1995; 95US-0468846.  
 XX (HUMA-) HUMAN GENOME SCI INC..  
 PA  
 PI Adams MD, Fuldner RA, Weissner PS, Wei YF;  
 XX WPI: 1997-043119/04.  
 DR N-PSDB; 747652.  
 XX  
 PT DNA encoding transforming growth factor alpha HII polypeptide  
 PT used to promote wound healing, to restore neurological functioning  
 PT after trauma or AIDS dementia  
 PS  
 PS Claim 1: Fig 1; 69pp; English.  
 CC The present sequence is that of human transforming growth factor (TGF)  
 CC alpha-HI. TGF-alpha-HI or its inhibitor can be used to treat a patient  
 CC to promote wound healing, to restore neurological functioning after  
 CC trauma or AIDS dementia, to treat ocular disorders, etc.. Disease or a  
 CC susceptibility to disease can be diagnosed by determining a mutation in  
 CC the nucleic acid sequence encoding TGF-alpha-HI. Receptors to TGF-alpha-  
 CC HI can be used to identify (ant)agonists.  
 XX  
 SQ Sequence 380 AA:  
 Query Match 80.4%; Score 37; DB 18; Length 380;  
 Best Local Similarity 75.0%; Pred. No. 1.1e+02;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 GGCVCVQDG 8  
 DB 85 ggvckedg 92  
 RESULT 8  
 Y79120  
 ID Y79120 standard; Peptide; 8 AA.  
 AC Y79120;  
 XX  
 DT 05-JUN-2000 (first entry)  
 XX  
 DE Peptide antagonist of zonulin.  
 XX  
 KW Zonulin; antagonist; zonula occludens toxin receptor;  
 KW blood-brain barrier; antiinflammatory; cerebroprotective;  
 KW neuroprotective; dermatological; antiulcer; antiviral;  
 KW antibacterial; cytostatic; anti-HIV; vulnery; antiallergic;  
 KW hypotensive; immunosuppressive; antiparasitic; vasotropic;  
 KW gastrointestinal inflammation; therapy.  
 XX  
 OS Synthetic.  
 XX  
 PN WO200007609-A1.  
 PD 17-FEB-2000.  
 XX  
 XX 28-JUL-1999; 99WO-US16683.  
 XX  
 PR 03-AUG-1998; 98US-0127815.  
 XX

PA (UYMA-) UNIV MARYLAND BALTIMORE.  
 XX  
 PI Fasano A;  
 XX  
 DR WPI: 2000-205565/18.  
 XX  
 PT New peptide antagonist of zonulin useful as antiinflammatory agent for  
 PT treating cerebral ischemia, stroke, cerebral edema, gastritis,  
 PT shigellosis, viral gastroenteritis, meningitis, encephalomyelitis  
 XX  
 XX Claim 1: Page 46; 69pp; English.  
 CC This present sequence is that of a peptide antagonist of zonulin  
 CC (Z), one of 25 such peptides (see Y79105-29) of the invention,  
 CC which bind to a zonula occludens toxin (ZOT) receptor, yet do not  
 CC physiologically modulate the opening of mammalian tight junctions  
 CC (TJ). The peptide antagonists are based on a common motif of ZOT  
 CC and human zonulin, which is believed to be critical for receptor  
 CC binding. They can be prepared by chemical synthesis or by use of  
 CC recombinant DNA techniques. The peptide antagonists are used as an  
 CC antiinflammatory agents in the treatment of gastrointestinal  
 CC inflammation, where they bind to the ZOT receptor in the intestine  
 CC and yet does not physiologically modulate the opening of TJ in the  
 CC intestine. Gastrointestinal inflammation conditions give rise to  
 CC increased intestinal permeability and the peptide is useful for  
 CC treating intestinal conditions that cause protein losing enteropathy  
 CC caused by infection, e.g. Clostridium difficile infection,  
 CC enterocolitis, shigellosis, viral gastroenteritis, parasite  
 CC infestation, bacterial overgrowth, whipple's disease, diseases with  
 CC mucosal erosion or ulcerations, e.g. gastritis, gastric cancer,  
 CC collagenous colitis, inflammatory bowel disease, diseases marked by  
 CC lymphatic obstruction, e.g. congenital intestinal lymphangiectasia,  
 CC sarcoidosis lymphoma, mesenteric tuberculosis, and after surgical  
 CC correction of congenital heart disease with Fontan's operation,  
 CC mucosal diseases without ulceration, e.g. Menetrier's disease,  
 CC coeliac disease, eosinophilic gastroenteritis, and immune diseases,  
 CC e.g. systemic lupus erythematosus or food allergies, primarily to  
 CC milk.  
 XX  
 SQ Sequence 8 AA:  
 Query Match 78.3%; Score 36; DB 21; Length 8;  
 Best Local Similarity 87.5%; Pred. No. 3.2e+05;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 GGCVCVQDG 8  
 DB 1 ggvlvqdg 8  
 RESULT 9  
 Y79121  
 ID Y79121 standard; Peptide; 8 AA.  
 AC Y79121;  
 XX  
 DT 05-JUN-2000 (first entry)  
 XX  
 DE Peptide antagonist of zonulin.  
 XX  
 KW Zonulin; antagonist; zonula occludens toxin receptor;  
 KW blood-brain barrier; antiinflammatory; cerebroprotective;  
 KW neuroprotective; dermatological; antiulcer; antiviral;  
 KW antibacterial; cytostatic; anti-HIV; vulnery; antiallergic;  
 KW hypotensive; immunosuppressive; antiparasitic; vasotropic;  
 KW gastrointestinal inflammation; therapy.  
 XX  
 OS Synthetic.  
 XX  
 PN WO200007609-A1.  
 PD 17-FEB-2000.  
 XX

XX 28-JUL-1999; 99WO-US16683.  
 XX 03-AUG-1998; 98US-0127815.  
 XX (UYMA-) UNIV MARYLAND BALTIMORE.  
 PA Fasano A;  
 PI  
 XX WPI; 2000-205565/18.  
 DR  
 XX  
 PT New peptide antagonist of zonulin useful as antiinflammatory agent for  
 PT treating cerebral ischemia, stroke, cerebral edema, gastritis,  
 PT shigellosis, viral gastroenteritis, meningitis, encephalomyelitis  
 XX  
 PS Claim 1; Page 46; 69pp; English.

CC This present sequence is that of a peptide antagonist of zonulin  
 CC (2), one of 25 such peptides (see Y79105-29) of the invention,  
 CC which bind to a zonula occludens toxin (ZOT) receptor, yet do not  
 CC physiologically modulate the opening of mammalian tight junctions  
 CC (17). The peptide antagonists are based on a common motif of ZOT  
 CC and human zonulins, which is believed to be critical for receptor  
 CC binding. They can be prepared by chemical synthesis or by use of  
 CC recombinant DNA techniques. The peptide antagonists are used as an  
 CC antiinflammatory agents in the treatment of gastrointestinal  
 CC inflammation, where they bind to the ZOT receptor in the intestine  
 CC and yet does not physiologically modulate the opening of tight  
 CC junctions. Gastrointestinal inflammation conditions give rise to  
 CC increased intestinal permeability and the peptide is useful for  
 CC treating intestinal conditions that cause protein losing enteropathy  
 CC caused by infection, e.g. Clostridium difficile infection,  
 CC enterocolitis, shigellosis, viral gastroenteritis, parasite  
 CC infection, bacterial overgrowth, whipple's disease, diseases with  
 CC mucosal erosion or ulcerations, e.g. gastritis, gastric cancer,  
 CC collagenous colitis, inflammatory bowel disease, diseases marked by  
 CC lymphatic obstruction, e.g. congenital intestinal lymphangiectasia,  
 CC sarcoidosis lymphoma, mesenteric tuberculosis, and after surgical  
 CC correction of congenital heart disease with Fontan's operation,  
 CC mucosal diseases without ulceration, e.g. Menetrier's disease,  
 CC coeliac disease, eosinophilic gastroenteritis, and immune diseases,  
 CC e.g. systemic lupus erythematosus or food allergies, primarily to  
 CC milk.

CC Sequence 8 AA;

Query Match 78.3%; Score 36; DB 21; Length 8;  
 Best Local Similarity 75.0%; Pred. No. 3.2e+05;

Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 GGVCVODG 8  
 11:1111  
 Db 1 gglcvpg 8

RESULT 10

Y42459 Y42459 standard; Protein; 115 AA.

AC Y42459;

DT 21-DEC-1999 (first entry)

DE Human guanylin preprohormone sequence.

XX Human; guanylin; preprohormone; intestinal guanylate cyclase regulator;  
 KW bacterium; heat stable enterotoxin; laxative; constipation.  
 XX  
 OS Homo sapiens.

PI Key Location/Qualifiers  
 FH Misc-difference 47  
 PT

FT /label= unknown  
 FT /note= "encoded by GAN"

XX US5969097-A.

XX 19-OCT-1999.

XX 23-JUN-1992; 92US-0903029.

XX 23-JUN-1992; 92US-0903029.

XX (SEAR ) SEARLE & CO G D.

XX Wiegand RC, Currie MG, Fok KF;

XX WPI; 1999-590442/50.

XX N-PDB; 222677.

PT Isolated protein used as a laxative in the treatment of constipation -  
 PS Disclosure; Fig 1; 14pp; English.

CC This sequence represents a human guanylin preprohormone which is  
 CC cleaved to generate a 15 amino acid mature peptide (Y42458). This  
 CC sequence is the encoded protein from a composite DNA sequence isolated  
 CC by probing a human duodenum cDNA library with the corresponding rat  
 CC coding sequence. Guanylin is an intestinal guanylate cyclase regulator  
 CC which has similarity to bacterial heat stable enterotoxins which  
 CC activate intestinal guanylate cyclases. The guanylin peptide can be  
 CC used as a laxative in the treatment of constipation.

CC Sequence 115 AA;

Query Match 78.3%; Score 36; DB 20; Length 115;  
 Best Local Similarity 87.5%; Pred. No. 52;

Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GGVCVODG 8  
 11111111  
 Db 20 ggvtyqdg 27

RESULT 11

Y42460 Y42460 standard; Protein; 115 AA.

AC Y42460;

DT 21-DEC-1999 (first entry)

DE Human guanylin preprohormone sequence.

XX Human; guanylin; preprohormone; intestinal guanylate cyclase regulator;  
 KW bacterium; heat stable enterotoxin; laxative; constipation.  
 XX

OS Homo sapiens.

XX US5969097-A.

XX 19-OCT-1999.

XX 23-JUN-1992; 92US-0903029.

XX 23-JUN-1992; 92US-0903029.

XX (SEAR ) SEARLE & CO G D.

XX Wiegand RC, Currie MG, Fok KF;

XX WPI; 1999-590442/50.

PT Isolated protein used as a laxative in the treatment of constipation -

XX PS Disclosure: Fig 2; 14pp; English.

CC CC This sequence represents a human guanylin preprohormone which is cleaved to generate a 15 amino acid mature peptide (Y42458). Guanylin is an intestinal guanylate cyclase regulator which has similarity to bacterial heat stable enterotoxins which activate intestinal guanylate cyclases.

CC CC The guanylin peptide can be used as a laxative in the treatment of constipation.

CC CC

XX SQ Sequence 115 AA;

Query Match 78.3%; Score 36; DB 20; Length 115;  
Best Local Similarity 87.5%; Pred. No. 52;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GGVCVODG 8  
1111111  
Db 20 ggvvtvgdg 27

RESULT 12

Y27892  
ID Y27892 standard; Protein: 115 AA.

XX AC Y27892;

XX DT 30-JUL-1999 (first entry)

XX DE Human secreted protein encoded by gene No. 116.

XX KW Human; secreted protein; fusion protein; gene therapy; protein therapy; diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia; developmental abnormality; foetal deficiency; blood; allergy; renal; immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma; inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS; cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus; osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion; endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.

XX OS Homo sapiens.

XX PN WO9924836-A1.

XX PD 20-MAY-1999.

XX PF 04-NOV-1998; 98WO-US23435.

XX PR 17-NOV-1997; 97US-0066100.  
PR 07-NOV-1997; 97US-0064900.  
PR 07-NOV-1997; 97US-0064908.  
PR 07-NOV-1997; 97US-0064911.  
PR 07-NOV-1997; 97US-0064912.  
PR 07-NOV-1997; 97US-0064983.  
PR 07-NOV-1997; 97US-0064984.  
PR 07-NOV-1997; 97US-0064985.  
PR 07-NOV-1997; 97US-0064987.  
PR 07-NOV-1997; 97US-0064988.  
PR 17-NOV-1997; 97US-0066090.  
PR 17-NOV-1997; 97US-0066094.  
PR 17-NOV-1997; 97US-0066095.  
PR 17-NOV-1997; 97US-0066089.

XX PA (HUMA-) HUMAN GENOME SCI INC.

XX PI Carter KC, Edner R, Moore GA, Feng P, Janat F;  
PI Kyaw H, Lafleur DW, Moore PA, Ni J, Olsen HS, Rosen CA;  
PI Ruben SM, Shi Y, Soppet DR, Wei Y;  
XX WPI: 1999-337740/28.  
XX N-PSDB: X85048.

PT PT New human secreted proteins and coding sequences useful for treating disorders of the immune system and hyperproliferative disorders

XX PS Disclosure: Page 166; 507pp; English.

XX CC This sequence represents a secreted human protein encoded by the gene clone detailed in the descriptor line. The gene can be used to generate fusion proteins by linking to the gene to a human immunoglobulin FC portion (e.g. X84924) for increasing the stability of the fused protein as compared to the human protein only.

CC CC The invention relates to 125 novel genes and their fragments (nucleic acid sequences: X84933-X85057; amino acid sequences Y27567-Y27933) which are useful for preventing, treating or ameliorating medical conditions e.g. by protein or gene therapy. Also, pathological conditions can be diagnosed by determining the amount of the new polypeptides in a sample or by determining the presence of mutations in the new polynucleotides. Specific uses are described for each of the 125 polynucleotides, based on CC which tissues they are most highly expressed in (see X84933 for described CC uses).

XX SQ Sequence 115 AA;

Query Match 78.3%; Score 36; DB 20; Length 115;  
Best Local Similarity 87.5%; Pred. No. 52;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GGVCVODG 8  
1111111  
Db 1 ggvsvvgdg 8

RESULT 13

W05520  
ID W05520 standard; Protein: 336 AA.

XX AC W05520;

XX DT 15-JAN-1997 (first entry)

XX DE HCMV Toledo strain UR151 protein (clone tol.22).

XX DE HCMV; HCMV; vaccine; diagnosis; UR151.

XX OS Human cytomegalovirus Toledo strain.

XX FH Key Location/Qualifiers

FT FT MISC-difference 47 /note= "unidentified amino acid"

FT FT MISC-difference 49 /note= "unidentified amino acid"

FT FT MISC-difference 115 /note= "unidentified amino acid"

XX PN W09630387-A1.

XX PD 03-OCT-1996.

XX PF 26-MAR-1996; 96WO-US04100.

XX PR 31-MAR-1995; 95US-0414926.

XX PA (AVIR-) AVIRON.

XX PI Cha T, Spaete R;

XX DR WPI: 1996-455265/45.  
XX N-PSDB: T41418.

XX PT New isolated human cytomegalovirus nucleic acid - from Towne and Toledo strains, used to develop prods. for the diagnosis, prevention and treatment of human CMV infections

PS Claim 5; Page 88-89; 150pp; English.

XX Novel protein UL151 (W05520) is the product of an open reading  
 CC frame found in a novel nucleic acid (T41418). Isolated from the  
 CC Toledo strain of human cytomegalovirus (HCMV). UL151 and other  
 CC novel (see also W05507-13) and known (see also W05500-01) proteins  
 CC from HCMV genome, as well as new proteins (see also W05521-24)  
 CC in the prodn. of subunit vaccines against HCMV. They may be  
 CC surface glycoproteins that are immunogenic or responsible for  
 CC tissue tropism, or may influence the immune response of an infected  
 CC individual.

SO Sequence 336 AA:

Query Match 78.3%; Score 36; DB 17; Length 336;  
 Best Local Similarity 75.0%; Pred. No. 1.4e+02;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGCVCVDG 8  
 Db 72 ggcvcgedg 79

RESULT 14  
 ID Y79110 standard; Peptide: 8 AA.  
 AC Y79110;  
 XX 05-JUN-2000 (first entry)  
 DE Peptide antagonist of zonulin.  
 XX Zonulin; antagonist; zonula occludens toxin receptor;  
 KW blood-brain barrier; antiinflammatory; cerebroprotective;  
 KW neuroprotective; dermatological; antiulcer; antiviral;  
 KW antibacterial; cytostatic; anti-HIV; vulnery; antiallergic;  
 KW hypotensive; immunosuppressive; antiparasitic; vasotropic;  
 KW gastrointestinal inflammation; therapy.  
 XX OS Synthetic.  
 XX WO200007609-A1.  
 PN 17-FEB-2000.  
 XX 28-JUL-1999; 99WO-US:6683.  
 PF 03-AUG-1998; 98US-01:7815.  
 PR (UYMA-) UNIV MARYLAND BALTIMORE.  
 PA Fasanio A;  
 PI WPI: 2000-205565/18.  
 DR New peptide antagonist of zonulin useful as antiinflammatory agent for  
 PT treating cerebral ischemia, stroke, cerebral edema, gastritis,  
 PT shigellosis, viral gastroenteritis, meningitis, encephalomyelitis -  
 XX Claim 1; Page 42; 69pp; English.

XX This present sequence is that of a peptide antagonist of zonulin  
 CC (2), one of 25 such peptides (see Y79105-29) of the invention,  
 CC which bind to a zonula occludens toxin (ZOT) receptor, yet do not  
 CC physiologically modulate the opening of mammalian tight junctions  
 CC (TJ). The peptide antagonists are based on a common motif of ZOT  
 CC and human zonulins, which is believed to be critical for receptor  
 CC binding. They can be prepared by chemical synthesis or by use of  
 CC recombinant DNA techniques. The peptide antagonists are used as an  
 CC antiinflammatory agents in the treatment of gastrointestinal

CC inflammation, where they bind to the ZOT receptor in the intestine  
 CC and yet does not physiologically modulate the opening of TJ in the  
 CC intestine. Gastrointestinal inflammation conditions give rise to  
 CC increased intestinal permeability and the peptide is useful for  
 CC treating intestinal conditions that cause protein losing enteropathy  
 CC caused by infection, e.g. Clostridium difficile infection,  
 CC enterocolitis, shigellosis, viral gastroenteritis, parasite  
 CC infection, bacterial overgrowth, whipple's disease, diseases with  
 CC mucosal erosion or ulcerations, e.g. gastritis, gastric cancer,  
 CC collagenous colitis, inflammatory bowel disease, diseases marked by  
 CC lymphatic obstruction, e.g. congenital intestinal lymphangiectasia,  
 CC sarcoidosis lymphoma, mesenteric tuberculosis, and after surgical  
 CC correction of congenital heart disease with Fontan's operation,  
 CC mucosal diseases without ulceration, e.g. Menetrier's disease,  
 CC coeliac disease, eosinophilic gastroenteritis, and immune diseases,  
 CC e.g. systemic lupus erythematosus or food allergies, primarily to  
 CC milk.

SO Sequence 8 AA:

Query Match 76.1%; Score 35; DB 21; Length 8;  
 Best Local Similarity 75.0%; Pred. No. 3.2e+05;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGCVCVDG 8  
 Db 1 ggcvcgdg 8

RESULT 15  
 ID R20006 standard; Protein: 399 AA.  
 AC R20006;  
 XX 31-MAR-1992 (first entry)  
 DE Zonula occludens toxin.  
 XX ZOT; cholera; vaccine; enterotoxin; diarrhoea.  
 KW Vibrio cholerae.  
 XX OS Vibrio cholerae.  
 XX WO9118979-A.  
 PN 12-DEC-1991.  
 XX 05-JUN-1991; 91WO-US03812.  
 PF 05-JUN-1990; 90US-0533315.  
 PR (UYMA-) UNIV MARYLAND BALTI.  
 PA Kaper JB, Baudry-Maurelli B, Fasanio A;  
 PI WPI: 1992-007465/01.  
 DR N-PSDB: Q20185.  
 PT New Vibrio cholerae strains - comprise restriction endonuclease  
 PT fragment encoding toxin, used as vaccines against cholera  
 XX Disclosure: Fig 18; 83pp; English.

XX The amino acid sequence is that of the zonula occludens toxin (ZOT).  
 CC It may be responsible for diarrhoea in some strains of cholera and  
 CC the ZOT gene or fragments of it are deleted from strains of Vibrio  
 CC cholerae (V.c.) to be used as vaccines. These V.c. strains have 100%  
 CC efficacy in protecting humans against subsequent infection with a  
 CC strain of a similar serotype and avoid undesirable side effects such  
 CC as diarrhoea, nausea and cramping. Cultures of these strains may be  
 CC used for prodn. of vaccines against cholera.

Sequence 399 AA:

Query Match 76.1%; Score 35; DB 13; Length 399;  
 Best Local Similarity 75.0%; Pred. No. 2.3e+02;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGVCVQDG 8  
 | :|||||

DB 291 grlcvcqdq 298

Search completed: June 13, 2001, 14:14:37  
 Job time: 378 sec

*This Page Blank (uspto)*

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 13, 2001, 14:10:45 ; Search time 130.61 Seconds

(without alignments)  
4.209 Million cell updates/sec

Title: PCT-US01-05825A-14

Perfect score: 46

Sequence: 1 GGCVCVDG 8

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 198801 seqs, 68722935 residues

Total number of hits satisfying chosen parameters: 198801

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR67:\*  
2: PIR1:\*  
3: PIR3:\*  
4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	41	89.1	196	2	F71525
2	41	89.1	196	2	H81681
3	41	89.1	196	2	G01639
4	36	78.3	115	1	A46279
5	36	78.3	172	2	S27022
6	36	78.3	266	2	S71025
7	36	78.3	633	2	T47524
8	36	78.3	680	2	T30620
9	36	78.3	713	2	I50128
10	36	78.3	822	2	S19947
11	36	78.3	822	2	B49151
12	35	76.1	399	2	B82197
13	35	76.1	399	2	A43864
14	35	76.1	402	2	C69110
15	35	76.1	491	2	D83353
16	35	76.1	1031	2	T06130
17	34.5	75.0	704	2	A48040
18	34	73.9	153	2	A27179
19	34	73.9	256	2	T16805
20	34	73.9	266	2	B82655
21	34	73.9	461	2	S69194
22	34	73.9	461	2	D70532
23	34	73.9	468	2	D70532
24	34	73.9	470	2	F36953
25	34	73.9	473	2	A56175
26	34	73.9	589	2	T08466
27	34	73.9	589	2	T08466
28	34	73.9	732	2	T01208
29	34	73.9	788	2	T07667

30	34	73.9	860	2	JC4939
31	34	73.9	861	2	A48825
32	34	73.9	876	2	A49508
33	34	73.9	910	2	A53137
34	34	73.9	913	2	A48280
35	34	73.9	1487	1	CGH06C
36	34	73.9	1492	2	A40333
37	34	73.9	1515	2	B46602
38	34	73.9	1548	2	S34583
39	34	73.9	2531	2	A46019
40	34	73.9	3670	2	T36249
41	33.5	72.8	391	2	T04486
42	33	71.7	206	2	S18250
43	33	71.7	213	2	A75471
44	33	71.7	282	2	F64091
45	33	71.7	440	2	T43198

#### ALIGNMENTS

beta-glucosidase (Notch homolog MOC protein-tyrosine k tyrosine kinase re receptor tyrosine collagen alpha 1(I) collagen alpha 1(I) glutamate synthase serine proteinase Notch-1 protein - CDA peptide synthe probable copropor collagen alpha 1(I) probable nitrogen lipopolysaccharide nitrilase/Fhl1 pro

RESULT 1  
F71525  
hypothetical protein CT349 - Chlamydia trachomatis (serotype D, strain UM3/Cx)  
C:Species: Chlamydia trachomatis  
C>Date: 13-Sep-1998 #sequence-revision 13-Sep-1998 #text-change 08-Oct-1999  
C:Accession: F71525  
R:Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitche  
A>Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia t  
A:Reference number: A71570; MUID:99000809  
A:Accession: F71525  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-196 <ARN>  
A:Cross-references: GB:AE001308; GB:AE001273; NID:g3328766; PIDN:AAC67944.1; PID:g332  
A:Experimental source: serotype D, strain UM-3/Cx  
C:Genetics:  
A:Gene: CT349  
C:Superfamily: septum formation protein mat

Query Match  
Best Local Similarity 89.1%; Score 41; DB 2; Length 196;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
OY 1 GGCVCVDG 8  
DB 150 GGCVCVDG 157

RESULT 2  
H81681  
mat protein TC0628 [imported] - Chlamydia muridarum (strain N19g)  
C:Species: Chlamydia muridarum, Chlamydia trachomatis MOPn  
C>Date: 31-Mar-2000 #sequence-revision 31-Mar-2000 #text-change 19-May-2000  
C:Accession: H81681  
R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hicke  
C.; Dodson, R.; Gwinn, M.; Nelson, W.; Nelson, R.; Kolonay, J.; McClarty, G.; Salzbe  
Nucleic Acids Res. 28, 1397-1406, 2000  
A>Title: Genome sequences of Chlamydia trachomatis MOPn and Chlamydia pneumoniae AR39  
A:Reference number: AB1500; MUID:20150255  
A:Accession: H81681  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-196 <TET>  
A:Cross-references: GB:AE002331; GB:AE002160; NID:g7190662; PIDN:AAF39457.1; PID:g719  
A:Experimental source: strain N19g (MOPn)  
C:Genetics:  
A:Gene: TC0628  
C:Superfamily: septum formation protein mat

Query Match 89.1%; Score 41; DB 2; Length 196;  
 Best Local Similarity 87.5%; Pred. No. 1.8;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 Oy 1 GGCVCODG 8  
 ||| |||  
 Db 150 GGCVCODG 157

RESULT 3  
 G01639  
 C:Species: Homo sapiens - human  
 C:Date: 21-Dec-1996 #sequence\_revision 06-Jun-1997 #text\_change 17-Nov-2000  
 C:Accession: G01639  
 R:ELD, D.W.  
 Submitted to the EMBL Data Library, January 1995  
 A:Reference number: G07997  
 A:Accession: G01639  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-380 <EID>  
 A:Cross-references: EMBL:U19878; NID:9755465; PIDN:AA6622.1; PID:9755466  
 C:Superfamily: unassigned IGF-related proteins; EGF homology  
 F:275-310/Domain: EGF homology <EGF>

Query Match 80.4%; Score 37; DB 2; Length 380;  
 Best Local Similarity 75.0%; Pred. No. 18;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Oy 1 GGCVCODG 8  
 ||| |||  
 Db 85 GGCVCODG 92

RESULT 4  
 A66279  
 C:Species: Homo sapiens (man)  
 C:Date: 21-Sep-1993 #sequence\_revision 26-May-1995 #text\_change 08-Dec-2000  
 C:Accession: A66279; S29228; S29807  
 R:de Sauvage, F.J.; Keshav, S.; Khang, W.J.; Gillett, N.; Henzel, W.; Goeddel, D.V.  
 Proc. Natl. Acad. Sci. U.S.A. 89, 9089-9093, 1992  
 A:Title: Precursor structure, expression, and tissue distribution of human guanylin.  
 A:Reference number: A66279; MUID:93028409  
 A:Accession: A66279  
 A:Molecule type: mRNA  
 A:Residues: 1-115 <DEI>  
 A:Cross-references: GB:M95174; NID:9306823; PIDN:AA58625.1; PID:9306824  
 R:Wiegand, R.C.; Kato, J.; Huang, M.D.; Fok, K.F.; Kachur, J.F.; Currie, M.G.  
 FEBS Lett. 311, 150-154, 1992  
 A:Title: Human guanylin: cDNA isolation, structure, and activity.  
 A:Reference number: S29228; MUID:93011964  
 A:Accession: S29228  
 A:Molecule type: mRNA  
 A:Residues: 1-115 <ME>  
 A:Cross-references: GB:M97496; NID:9183414; PIDN:AA55915.1; PID:9183415  
 R:Kuhn, M.; Reide, M.; Adenmann, K.; Schulz-Knappe, P.; Gerzer, R.; Helm, J.M.; Forssman  
 FEBS Lett. 318, 205-209, 1993  
 A:Title: The circulating bioactive form of human guanylin is a high molecular weight peptide.  
 A:Reference number: S29807; MUID:93178628  
 A:Accession: S29807  
 A:Molecule type: protein  
 A:Residues: 22-68 <KUH>  
 A:Experimental source: Plasma  
 C:Note: amino-terminal sequencing of mature form and molecular weight of mature form by n of the same receptor.  
 C:Genetics:  
 A:Gene: GDB:GUC2  
 A:Cross-references: GDB:136450; OMIM:139392

A:Map position: 1035-1p34  
 C:Superfamily: guanylin  
 C:Keywords: hormone; intestine  
 F:1-21/Domain: signal sequence #status predicted <SIG>  
 F:22-115/Product: guanylin #status experimental <MNT>

Query Match 78.3%; Score 36; DB 1; Length 115;  
 Best Local Similarity 87.5%; Pred. No. 9.3;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 GGCVCODG 8  
 ||| |||  
 Db 20 GGCVCODG 27

RESULT 5  
 S27022  
 fibroblast growth factor receptor - Japanese medaka (fragment)  
 N:Contans: protein-tyrosine kinase (EC 2.7.1.112)  
 C:Species: Oryzias latipes (Japanese medaka)  
 C:Date: 19-Mar-1997 #sequence\_revision 24-Jul-1997 #text\_change 20-Jun-2000  
 C:Accession: S27022  
 R:Emori, Y.; Yasuoka, A.; Saigo, K.  
 FEBS Lett. 314, 176-178, 1992  
 A:Title: Identification of four FGF receptor genes in Medaka fish (Oryzias latipes).  
 A:Reference number: S27019; MUID:93093167  
 A:Accession: S27022  
 A:Molecule type: not compared with conceptual translation  
 A:Status: not compared with conceptual translation  
 A:Molecule type: mRNA  
 A:Residues: 1-172 <EMO>  
 A:Cross-references: GB:D13553; NID:9222938; PIDN:BA02752.1; PID:9222939  
 C:Keywords: alternative splicing; ATP; autophosphorylation; growth factor receptor; pr  
 F:1-172/Domain: protein kinase homology (fragment) <KIN>  
 F:23/40,132/Active site: Lys, Glu, Asp #status predicted  
 F:137,150/Binding site: magnesium (Asn, Asp) #status predicted  
 F:163/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation) #status predi

Query Match 78.3%; Score 36; DB 2; Length 172;

Best Local Similarity 85.7%; Pred. No. 13;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 2 GGCVCODG 8  
 ||| |||  
 Db 58 GGCVCODG 64

RESULT 6  
 S71025  
 lipopolysaccharide biosynthesis protein lfc2B - Haemophilus influenzae (strain RM7004  
 C:Species: Haemophilus influenzae  
 C:Date: 15-Feb-1997 #sequence\_revision 13-Mar-1997 #text\_change 20-Jun-2000  
 C:Accession: S71025  
 R:High, N.J.; Jennings, M.P.; Moxon, E.R.  
 Mol. Microbiol. 20, 165-174, 1996  
 A:Title: Random repeats of the tetramer 5'-CAAT-3' present in lfc2A are required for  
 A:Reference number: S71024; MUID:97014379  
 A:Accession: S71025  
 A:Molecule type: DNA  
 A:Residues: 1-266 <HIG>  
 A:Cross-references: EMBL:Z54182; NID:91150402; PIDN:CA90892.1; PID:91150403  
 A:Experimental source: strain RM7004  
 C:Genetics:  
 A:Gene: lfc2B  
 C:Function:  
 A:Description: required for the biosynthesis of a phase-variable lipopolysaccharide s  
 C:Superfamily: lipopolysaccharide biosynthesis-associated protein

Query Match 78.3%; Score 36; DB 2; Length 266;  
 Best Local Similarity 85.7%; Pred. No. 20;



Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 GVCVODG 8  
 |||||:1  
 Db 199 GVCVODG 205

RESULT 7  
 T47524  
 hypothetical protein F16L2.50 - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 20-Apr-2000 #sequence\_revision 20-Apr-2000 #text\_change 20-Apr-2000  
 C:Accession: T47524  
 R:Jordan, N.; Bangerf, S.; Wiedemann, R.; Voss, H.; Unseld, M.; Mewes, H.W.; Rudd, S.;  
 submitted to the Protein Sequence Database, March 2000  
 A:Reference number: 224468  
 A:Accession: T47524  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-633 <JOB>  
 A:Cross-references: EMBL:AL162459  
 A:Experimental source: cultivar Columbia; BAC clone F16L2  
 C:Genetics:  
 A:Map position: 3  
 A:Introns: 233/3: 257/2: 582/2: 621/3  
 A:Note: F16L2.50

Query Match  
 Best Local Similarity 78.3%; Score 36; DB 2; Length 633;  
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 GVCVODG 8  
 |||||  
 Db 624 GGCCTDDG 631

RESULT 8  
 T30620  
 hypothetical protein 18L - Moluscum contagiosum virus 1  
 N:Alternate names: MC018L  
 C:Species: Moluscum contagiosum virus 1  
 C:Date: 05-Nov-1999 #sequence\_revision 05-Nov-1999 #text\_change 11-May-2000  
 C:Accession: T30620  
 R:Senkevich, T.G.; Bugert, J.J.; Sisler, J.R.; Koonin, E.V.; Darai, G.; Moss, B.  
 Science 273, 813-816, 1996  
 A:Title: Genome sequence of a human tumorigenic poxvirus: Prediction of specific host re  
 A:Reference number: 220876; MUID:96325459  
 A:Accession: T30620  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-680 <SEN>  
 A:Cross-references: EMBL:U60315; PIDN:AC55146.1  
 C:Genetics:  
 A:Note: MC018L

Query Match  
 Best Local Similarity 78.3%; Score 36; DB 2; Length 680;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 GVCVODG 8  
 |||||  
 Db 151 GACVODG 157

RESULT 9  
 T50128  
 fibroblast growth factor receptor - quail  
 C:Species: Coturnix coturnix (quail)  
 C:Date: 13-Sep-1996 #sequence\_revision 13-Sep-1996 #text\_change 16-Jul-1999  
 C:Accession: T50128; S42803  
 R:Marcelle, C.; Elchmann, A.; Haley, O.; Breant, C.; Le Douarin, N.M.

Development 120, 683-694, 1994  
 A:Title: Distinct developmental expression of a new avian fibroblast growth factor re  
 A:Reference number: T50128; MUID:94215505  
 A:Accession: T50128  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-713 <MAR>  
 A:Cross-references: EMBL:X76885; NID:9440139; PIDN:CAA54213.1; PID:9440140  
 C:Genetics:  
 A:Gene: FREK  
 C:Superfamily: basic fibroblast growth factor receptor 1; immunoglobulin homology; pr  
 C:Keywords: ATP; growth factor receptor  
 F:176-247/Domain: immunoglobulin homology <IMM>  
 F:377-662/Domain: protein kinase homology <KIN>  
 F:385-393/Region: protein kinase ATP-binding motif

Query Match  
 Best Local Similarity 78.3%; Score 36; DB 2; Length 713;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 GVCVODG 8  
 |||||  
 Db 450 GVCCTDDG 456

RESULT 10  
 S19947  
 fibroblast growth factor receptor - Iberian ribbed newt  
 C:Species: Pleurodeles waltl (Iberian ribbed newt)  
 C:Date: 06-Jan-1995 #sequence\_revision 06-Jan-1995 #text\_change 16-Jul-1999  
 C:Accession: S19947  
 R:Shi, D.L.; Feige, J.J.; Riou, J.F.; Desimone, D.W.; Boucaut, J.C.  
 submitted to the EMBL Data Library, March 1992  
 A:Description: Receptors during early development of the urodele Pleurodeles waltl  
 A:Reference number: S19947  
 A:Accession: S19947  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-822 <SHI>  
 A:Cross-references: EMBL:X65059  
 C:Superfamily: basic fibroblast growth factor receptor 1; immunoglobulin homology; pr  
 C:Keywords: ATP; growth factor receptor  
 F:283-354/Domain: immunoglobulin homology <IMM>  
 F:488-769/Domain: protein kinase homology <KIN>  
 F:492-500/Region: protein kinase ATP-binding motif

Query Match  
 Best Local Similarity 78.3%; Score 36; DB 2; Length 822;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 GVCVODG 8  
 |||||  
 Db 557 GVCCTDDG 563

RESULT 11  
 B49151  
 fibroblast growth factor receptor 4 - Iberian ribbed newt  
 C:Species: Pleurodeles waltl (Iberian ribbed newt)  
 C:Date: 19-Dec-1993 #sequence\_revision 18-Nov-1994 #text\_change 16-Jul-1999  
 C:Accession: B49151  
 R:Shi, D.L.; Feige, J.J.; Riou, J.F.; Desimone, D.W.; Boucaut, J.C.  
 Development 116, 261-273, 1992  
 A:Title: Differential expression and regulation of two distinct fibroblast growth fac  
 A:Reference number: A49151; MUID:93130775  
 A:Accession: B49151  
 A:Status: preliminary  
 A:Molecule type: nucleic acid  
 A:Residues: 1-822 <SHI>  
 A:Cross-references: GB:X65059; NID:964252; PIDN:CAA46192.1; PID:964253  
 A:Experimental source: tail-bud

A:Note: sequence extracted from NCBI backbone (NCBIN:122598, NCBI:122599)  
C:Superfamily: basic fibroblast growth factor receptor 1; immunoglobulin homology; prote  
C:Keywords: ATP; growth factor receptor  
F:283-354/Domain: immunoglobulin homology <IMM>  
F:484-769/Domain: protein kinase homology <KIN>  
F:492-500/Region: protein kinase ATP-binding motif

Query Match 78.3%; Score 36; DB 2; Length 822;  
Best Local Similarity 85.7%; Pred. No. 56;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 GGCYVDG 8  
DB 557 GGCYVDG 563

RESULT 12  
B82197  
zona occludens toxin VC1458 [Imported] - Vibrio cholerae (group O1 strain N16961)

C:Species: Vibrio cholerae  
C:Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 15-Sep-2000  
C:Accession: B82197  
R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwin, M.L.; Dodson, R.J.;  
Chadson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragol, I.; Sellers, F.  
I.; R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.  
Nucleotide 406, 477-483, 2000  
A:Title: DNA sequence of both chromosomes of the cholera pathogen Vibrio cholerae.  
A:Reference number: A82035; MUID:20406833  
A:Accession: B82197  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-399 <HEI>  
A:Cross-references: GB:AE004224; GB:AE003852; NID:g9655592; PIDN:AAF94615.1; GSPDB:GN001  
A:Experimental source: serogroup O1, strain N16961; biotype El Tor  
C:Genetics:  
A:Gene: VC1458  
A:Map position: 1

Query Match 76.1%; Score 35; DB 2; Length 399;  
Best Local Similarity 75.0%; Pred. No. 44;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 GGCYVDG 8  
DB 291 GRLCYVDG 298

RESULT 13

A43864  
zonula occludens toxin - Vibrio cholerae

C:Species: Vibrio cholerae  
C:Date: 10-Mar-1993 #sequence\_revision 18-Nov-1994 #text\_change 08-Oct-1999  
C:Accession: A43864  
R:Baudry, B.; Pasano, A.; Kelley, J.; Kapet, J.B.  
Infect. Immun. 60, 428-434, 1992  
A:Title: Cloning of a gene (zot) encoding a new toxin produced by Vibrio cholerae.  
A:Reference number: A43864; MUID:92112300  
A:Accession: A43864  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-399 <BAU>  
A:Cross-references: GB:M8356; NID:g155314; PIDN:AAA27582.1; PID:g155315  
A:Note: sequence extracted from NCBI backbone (NCBIN:77488, NCBI:77491)

Query Match 76.1%; Score 35; DB 2; Length 399;  
Best Local Similarity 75.0%; Pred. No. 44;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 GGCYVDG 8  
DB 134 GGCYVDG 141

DB 291 GRLCYVDG 298

RESULT 14

C69110

glutamate N-acetyltransferase - Methanobacterium thermoautotrophicum (strain Delta H)

C:Species: Methanobacterium thermoautotrophicum

C:Date: 03-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 29-Sep-1999

C:Accession: C69110

R:Smith, D.R.; Doucette-Stamm, L.A.; Delonghery, C.; Lee, H.; Dubois, J.; Aldredge, T.  
K.; S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.  
J. Bacteriol. 179, 7135-7155, 1997  
A:Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: fu

A:Reference number: A69000; MUID:98037514  
A:Accession: C69110  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-402 <MTB>

A:Cross-references: GB:AE000806; GB:AE000666; NID:g2621222; PIDN:AB84688.1; PID:g262  
C:Genetics:  
A:Gene: MTB182  
A:Start codon: GTG  
C:Superfamily: glutamate N-acetyltransferase arg

Query Match 76.1%; Score 35; DB 2; Length 402;  
Best Local Similarity 75.0%; Pred. No. 45;  
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 GGCYVDG 8  
DB 11 GGCYVDG 18

RESULT 15

D83353

mannitol dehydrogenase PA2342 [Imported] - Pseudomonas aeruginosa (strain PA01)

C:Species: Pseudomonas aeruginosa

C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000

C:Accession: D83353

R:Stover, C.K.; Pham, X.Q.; Ervin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.;  
Adman, S.; Yun, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lartig, K.; L.  
Nure, S.; Olson, M.V.  
Nucleotide 406, 959-964, 2000  
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa

A:Reference number: A82950; MUID:20437337  
A:Accession: D83353  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-491 <STO>

A:Cross-references: GB:AE004660; GB:AE004091; NID:g9948372; PIDN:AA05730.1; GSPDB:GN  
A:Experimental source: strain PA01  
C:Genetics:  
A:Gene: mtd; PA2342  
C:Superfamily: conserved hypothetical protein YEL070w

Query Match 76.1%; Score 35; DB 2; Length 491;  
Best Local Similarity 62.5%; Pred. No. 54;  
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 GGCYVDG 8  
DB 134 GGCYVDG 141

Search completed: June 13, 2001, 14:10:45  
Job time: 146 sec



*This Page Blank (uspto)*

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 13, 2001, 14:21:46 ; Search time 74.44 Seconds  
(without alignments)  
3.681 Million cell updates/sec

Title: PCT-US01-05825A-14  
Perfect score: 46  
Sequence: 1 GGVGVQDGC 8

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues  
Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries.

Database : SWISSPROT\_39:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	36	78.3	115	1	GUAN_HUMAN
2	36	78.3	266	1	L12B_HAEMIN
3	35	76.1	399	1	ZOT_VIBCH
4	34.5	75.0	704	1	MEPB_MOUSE
5	34	73.9	340	1	ASPG_FLAME
6	34	73.9	468	1	STHA_MYCTU
7	34	73.9	473	1	FE2_MYTGA
8	34	73.9	589	1	CAH_DUNSA
9	34	73.9	788	1	UGS3_SOLTU
10	34	73.9	860	1	BGL1_ASFAC
11	34	73.9	867	1	EP2_ELAHO
12	34	73.9	910	1	DDR1_RAT
13	34	73.9	911	1	DDR1_MOUSE
14	34	73.9	913	1	DDR1_HUMAN
15	34	73.9	1453	1	CA11_CHICK
16	34	73.9	1515	1	G1TB_AZOBH
17	34	73.9	1877	1	RCK5_MOUSE
18	34	73.9	2531	1	NTC1_MOUSE
19	33.5	72.8	391	1	HEM6_HORVU
20	33	71.7	457	1	Y765_HAEMIN
21	33	71.7	457	1	PH4H_CAEEL
22	33	71.7	512	1	AGT2_RAT
23	33	71.7	587	1	URE1_CLOPE
24	33	71.7	752	1	UGS3_PEA
25	33	71.7	788	1	TRSI_HCVMA
26	33	71.7	825	1	BGLS_HANAN
27	33	71.7	846	1	IRSI_HCVMA
28	33	71.7	979	1	PTPN_BOVIN
29	33	71.7	979	1	PTPN_HUMAN
30	33	71.7	979	1	PTPN_MOUSE
31	33	71.7	983	1	PTPN_RAT
32	33	71.7	1093	1	SM5B_MOUSE
33	33	71.7	1120	1	DPOL_RCMVM

34	33	71.7	1153	1	JAK1_MOUSE	P52332	mus musculus
35	33	71.7	1569	1	GLI3_XENLA	O91660	xenopus lae
36	33	71.7	2318	1	NTC3_MOUSE	O61982	mus musculus
37	32	69.6	109	1	GUAN_PIG	P79897	sus scrofa
38	32	69.6	316	1	CH18_POPTR	P16061	populus tri
39	32	69.6	491	1	TRF3_CHICK	O90643	gallus gall
40	32	69.6	515	1	CP11_MICRO	O92148	microgadus
41	32	69.6	571	1	APB3_MOUSE	O88888	mus musculus
42	32	69.6	575	1	APB3_HUMAN	O96018	homo sapien
43	32	69.6	589	1	DLI3_RAT	O88671	rattus norv
44	32	69.6	592	1	DLI3_MOUSE	O88516	mus musculus
45	32	69.6	618	1	DLI3_HUMAN	Q9nyj7	homo sapien

## ALIGNMENTS

RESULT	ID	GUAN_HUMAN	STANDARD:	PRT:	115 AA.
AC	002747;				
DT	01-JUL-1993 (Rel. 26, Created)				
DT	01-JUL-1993 (Rel. 26, Last sequence update)				
DT	01-OCT-2000 (Rel. 40, Last annotation update)				
DE	GUANYLIN PRECURSOR (GUANYLATE CYCLASE ACTIVATOR 2A).				
GN	GUCA2A OR GUCA2.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.				
OX	NCBI_TaxID=9606;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=Dundenum;				
RX	MEDLINE=93011964; PubMed=1327879;				
RA	Wiegand R.C., Kato J., Huang M.D., Fok K.F., Kachur J.F.,				
RA	Currie M.G.;				
RT	"Human guanylin: cDNA isolation, structure, and activity."				
RL	FEBS Lett. 311:150-154(1992).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=ileum;				
RX	MEDLINE=93028409; PubMed=1409606;				
RA	de Sauvage F.J., Keshav S., Kuang W.J., Gillett N., Henzel W.,				
RA	Goeddel D.V.;				
RT	"Precursor structure, expression, and tissue distribution of human				
RL	guanylin."				
RN	Proc. Natl. Acad. Sci. U.S.A. 89:9089-9093(1992).				
RP	[3]				
RX	SEQUENCE OF 22-68.				
RP	MEDLINE=93178628; PubMed=8095028;				
RA	Kuhn M., Ralda M., Adermann K., Schulz-Knappe P., Gerzer R.,				
RA	Heim J.-M., Forssmann W.-G.;				
RT	"The circulating bioactive form of human guanylin is a high molecular				
RL	weight peptide (10.3 kDa)."				
RN	FEBS Lett. 318:205-209(1993).				
RP	[4]				
RX	STRUCTURE BY NMR OF 101-115.				
RP	MEDLINE=95034794; PubMed=7947768;				
RA	Skellton N.J., Garcia K.C., Goeddel D.V., Quan C., Burnier J.P.;				
RT	"determination of the solution structure of the peptide hormone				
RL	guanylin: observation of a novel form of topological				
RT	stereoisomerism."				
RL	Biochemistry 33:13581-13592(1994).				
CC	-1- FUNCTION: ENDOGENOUS ACTIVATOR OF INTESTINAL GUANYLATE CYCLASE.				
CC	IF STIMULATES THIS ENZYME THROUGH THE SAME RECEPTOR BINDING REGION				
CC	AS THE HEAT-STABLE ENTEROTOXINS.				
CC	-1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN ILEUM AND COLON.				
CC	-1- SIMILARITY: BELONGS TO THE GUANYLIN FAMILY. ALSO SIMILAR TO HEAT-				
CC	STABLE ENTEROTOXINS.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				

CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: M91496; AAA5915.1; -  
DR EMBL: M9174; AAA58625.1; -  
DR PIR: A46279; A46279.  
DR PIR: S29228; S29228.  
DR PDB: 1GNA; 30-SEP-94.  
DR PDB: 1GNB; 30-SEP-94.  
DR MIM: 139392; -  
DR InterPro: IPR000879; -  
DR Pfam: PF02058; GuanYlin.1.  
DR PRINTS: PR00774; GUANYLIN.  
KW signal; 3d-structure.  
FT SIGNAL 1 21  
FT CHAIN 22 115 HMM-GUANYLIN.  
FT PEPTIDE 101 115  
FT DISULFID 104 112 GUANYLIN.  
FT DISULFID 107 115  
SQ SEQUENCE 115 AA; 12448 MW; C644C03BAFC26FA CRC64;

Query Match 73.3%; Score 36; DB 1; Length 115;  
Best Local Similarity 87.5%; Pred. NO. 5.3;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 GVCVQDG 8  
1111111  
Db 20 GGVTVQDG 27

RESULT 2  
L12B\_HAETN STANDARD; PRT; 266 AA.  
ID L12B\_HAETN STANDARD; PRT; 266 AA.  
AC 05394;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE LIPOOLIGOSACCHARIDE BIOSYNTHESIS PROTEIN LIC2B.  
GN LIC2B.  
OS Haemophilus influenzae.  
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
OC Haemophilus.  
OX NCBI\_TaxID=727;  
[1]  
SEQUENCE FROM N.A.  
RP STRAIN-BM 7004 / SEROTYPE B;  
RA High N.J., Jennings M.P., Moxon R.E.;  
RT Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: INVOLVED IN EXTRACELLULAR LIPOOLIGOSACCHARIDES (LOS)  
CC OF BIOSYNTHESIS AND VIRULENCE EXPRESSION. INVOLVED IN THE SYNTHESIS  
CC OF THE OLIGOSACCHARIDE MOIETY OF THE LOS MOLECULE BY ADDING GALAC  
CC (BY SIMILARITY).  
CC -1- SIMILARITY: BELONGS TO THE LEXA/PRINCE FAMILY OF  
CC GLYCOSYLTRANSFERASES.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

Query Match 78.3%; Score 36; DB 1; Length 266;  
Best Local Similarity 85.7%; Pred. NO. 12;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 GVCVQDG 8  
1111111  
Db 199 GVCVQDG 205

RESULT 3  
ZOT\_VIBCH STANDARD; PRT; 399 AA.  
ID ZOT\_VIBCH STANDARD; PRT; 399 AA.  
AC P38442; Q9L7Q6; Q9R3V6;  
DT 01-OCT-1994 (Rel. 30, Created)  
DT 01-OCT-2000 (Rel. 40, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE ZONA OCCUDENS TOXIN (ZONULAR OCCUDENS TOXIN).  
GN ZOT OR VC1458.  
OS Vibrio cholerae.  
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.  
OX NCBI\_TaxID=666;  
[1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-CLASSICAL; INABA 5698;  
RX MEDLINE=92112300; PubMed=1730472;  
RA Baudry B., Fasano A., Kelley J., Kaper J.B.;  
RT "Cloning of a gene (zot) encoding a new toxin produced by Vibrio  
RT cholerae.";  
RT Infect. Immun. 60:428-434(1992).  
[2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-KNIH002;  
RA Shin H.J., Park Y.C., Kim Y.C.;  
RT "Cloning and nucleotide sequence analysis of the virulence gene  
RT cassette from Vibrio cholerae KNIH002 isolated in Korea.";  
RT Misalimurrah Holji 35:205-210(1999).  
[3]  
RP SEQUENCE FROM N.A.  
RC STRAIN-O139-TOR OGAMA;  
RA Zhi-Yong H., Wei-Jie Z., Xiang-Fu W.;  
RT "Cloning and Expression of zot gene from Vibrio cholerae.";  
RT Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.  
[4]  
RP SEQUENCE FROM N.A.  
RC STRAIN-EL TOR 86015 / SEROTYPE O1;  
RA Kan B., Liu Y.O., Qi G.M., Gao S.Y.;  
RT Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.  
[5]  
RP SEQUENCE FROM N.A.  
RC STRAIN-EL TOR N16961 / SEROTYPE O1;  
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwin M.L.,  
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,  
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,  
RA McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,  
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,  
RA Fraser C.M.;  
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio  
RT cholerae.";  
RT Nature 406:477-483(2000).  
[6]  
RP CHARACTERIZATION.  
RX MEDLINE=91271365; PubMed=2052603;  
RA Fasano A., Baudry B., Pumpilin D.W., Wasserman S.S., Tall B.D.,  
RA Kelley J.M., Kaper J.B.;  
RT "Vibrio cholerae produces a second enterotoxin, which affects  
RT intestinal tight junctions.";  
RT Proc. Natl. Acad. Sci. U.S.A. 88:5242-5246(1991).  
CC -1- FUNCTION: INCREASES THE PERMEABILITY OF THE SMALL INTESTINE MUCOSA  
CC BY AFFECTING THE STRUCTURE OF INTERCELLULAR TIGHT JUNCTIONS  
CC (ZONULA OCCUDENS).

-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
-----  
DR EMBL; M83563; AAA27582.1; -  
DR EMBL; AF175708; AAD51358.1; -  
DR EMBL; AF123049; AAD26854.1; -  
DR EMBL; AF220606; AAD29547.1; -  
DR EMBL; AF004224; AAF94615.1; -  
DR PIR; A43864; A43864.  
DR TIGR; VC1458; -  
KW Enterotoxin; Toxin.  
FT VARIANT 45 45 M -> I (IN STRAIN 569B).  
FT VARIANT 100 100 V -> A (IN STRAINS 569B AND 86015).  
FT VARIANT 272 272 V -> A (IN STRAIN 569B).  
FT VARIANT 281 281 V -> A (IN STRAIN 569B).  
FT VARIANT 349 349 A -> S (IN STRAIN 86015).  
FT VARIANT 381 381 K -> R (IN STRAIN 86015).  
FT CONFLICT 386 399 IKTEKDKKGLNSIF -> VKKEKESEIIRKFL (IN REF.  
FT SEQUENCE 399 AA; 44903 MW; 3C7424B758176774 CRC64;  
SO  
Query Match Best Local Similarity 76.18; Score 35; DB 1; Length 399;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
OY 1 GGCVCVODG 8  
DB 291 GRLCVDG 298  
RESULT 4  
MEPR\_MOUSE  
ID MEPR\_MOUSE STANDARD; PRT; 704 AA.  
AC Q61847;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE MEPRIN A BETA-SUBUNIT PRECURSOR (EC 3.4.24.18) (ENDOPEPTIDASE-2).  
CN MEPR1B OR MEPR1B.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
OX NCBI\_Taxid=10090;  
[1]  
RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.  
RC TISSUE-Kidney;  
RX MEDLINE=94012651; PubMed=8407940;  
RA Gorbea C.M., Marchand P., Jiang W., Copeland N.G., Gilbert D.J.,  
RT Jenkins N.A., Bond J.S.;  
RT "Cloning, expression, and chromosomal localization of the mouse  
RT meprin beta subunit."  
RL J. Biol. Chem. 268:21035-21043(1993).  
RN [2]  
RP SEQUENCE FROM N.A. (ISOFORM BETA').  
RC TISSUE-Kidney;  
RX MEDLINE=96147211; PubMed=8567689;  
RA Dietrich J.M., Bond J.S., Jiang W.;  
RT "A novel meprin beta' mRNA in mouse embryonal and human colon  
RT carcinoma cells."  
RL J. Biol. Chem. 271:2271-2278(1996).  
CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF PROTEIN AND PEPTIDE SUBSTRATES  
CC PREFERENTIALLY ON CARBOXYL SIDE OF HYDROPHOBIC RESIDUES.  
CC -1- COFACTOR: BINDS ONE ZINC ION.  
CC -1- SUBUNIT: HETEROPTETRAMER OF TWO ALPHA AND TWO BETA SUBUNITS WHICH  
CC IS FORMED BY THE NON-COVALENT ASSOCIATION OF TWO DISULFIDE-LINKED  
CC HETERODIMERS.

-----  
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.  
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS, THE BETA AND BETA' SUBUNITS, AND  
CC DIFFER IN THEIR N-TERMINUS DUE TO DIFFERENTIAL PROMOTER USAGE AND  
CC ALTERNATIVE SPLICING.  
CC -1- TISSUE SPECIFICITY: THE BETA-SUBUNIT IS EXPRESSED IN KIDNEY,  
CC INTESTINAL BRUSH BORDERS, AND SALIVARY DUCTS. THE BETA'-ISOFORM  
CC HAS BEEN FOUND IN CARCINOMA CELLS.  
CC -1- INDUCTION: THE BETA'-SUBUNIT IS INDUCED BY THE MORPHOGEN RETINOIDS  
CC ACID.  
CC -1- PM: THIS PROTEIN UNDERGOES PROTEOLYTIC PROCESSING. BOTH FORMS  
CC ARE GLYCOSYLATED.  
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12A (ZINC  
CC METALLOPROTEASE); ALSO KNOWN AS THE ASTRACIN SUBFAMILY..  
CC -1- SIMILARITY: CONTAINS 1 MAM DOMAIN.  
CC -1- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
-----  
DR EMBL; L15193; AAA75234.1; -  
DR HSSP; P28825; IIAF.  
DR MEROPS; M12.002; -  
DR MEROPS; M12.004; -  
DR MGD; MGI:96964; Mep1b.  
DR InterPro; IPR000130; -  
DR InterPro; IPR000561; -  
DR InterPro; IPR000998; -  
DR InterPro; IPR001506; -  
DR InterPro; IPR02083; -  
DR Pfam; PF01400; Astacin; 1.  
DR Pfam; PF00008; EGF; 1.  
DR Pfam; PF00629; MAM; 1.  
DR Pfam; PF00917; MATN; 1.  
DR PRINTS; PR00020; MAMDOMAIN.  
DR PRINTS; PR00480; ASTACIN.  
DR PROSITE; PS00142; ZINC\_PROTEASE; 1.  
DR PROSITE; PS00740; MAM\_1; 1.  
DR PROSITE; PS00620; MAM\_2; 1.  
DR PROSITE; PS00022; EGF\_1; FALSE NEG.  
DR PROSITE; PS01186; EGF\_2; FALSE NEG.  
KW Transmembrane; Hydrolyase; Metalloprotease; Zinc; Glycoprotein;  
KW Zymogen; Signal; EGF-like domain; Alternative splicing.  
FT SIGNAL 1 20  
FT PROPEP 21 64  
FT CHAIN 65 704  
FT DOMAIN 21 654  
FT TRANSMEM 655 678  
FT DOMAIN 679 704  
FT DOMAIN 63 260  
FT DOMAIN 261 430  
FT DOMAIN 607 647  
FT METAL 153 153  
FT METAL 154 154  
FT ACT\_SITE 157 157  
FT METAL 163 163  
FT DISULFD 611 622  
FT DISULFD 616 631  
FT DISULFD 633 646  
FT CARBOHYD 193 193  
FT CARBOHYD 219 219  
FT CARBOHYD 255 255  
FT CARBOHYD 316 316  
FT CARBOHYD 422 422  
FT CARBOHYD 437 437  
FT CARBOHYD 529 529  
FT CARBOHYD 548 548  
FT CARBOHYD 593 593  
FT VASPLIC 1 27  
-----  
MDARHPWFLVFAIFLASGLPAPKPF -> MNSIAGPASR

FT SRHSEKRMKLLKAPRDGKMYMTFG (IN ISOFORM  
 FT BETA').  
 SO SEQUENCE 704 AA; 79548 MW; 2D610FDD74650F70 CRC64;

Query Match 75.0%; Score 34.5; DB 1; Length 704;  
 Best Local Similarity 77.8%; Pred. No. 52;  
 Matches 7; Conservative 1; Mismatches 0; Indels 1; Gaps 1;

OY 1 GGYC-VODG 8  
 DB 619 GGICVODG 627

RESULT 5  
 ID ASPG\_FLAME STANDARD; PRT; 340 AA.  
 AC 047898;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE N4-(BETA-N-ACETYLGLUCOSAMINYL)-L-ASPARAGINASE PRECURSOR (EC 3.5.1.26)  
 DE (GLYCOSYLASPARAGINASE) (ASPARTYLGLUCOSAMINIDASE) (N4-(N-ACETYL-BETA-  
 DE GLUCOSAMINYL)-L-ASPARAGINE AMIDASE) (AGA).  
 OS Flavobacterium meningosepticum.  
 OC Bacteria; CF8 group; Flavobacteriia; Flavobacteriaceae;  
 OC Chryseobacterium.  
 OX NCBI\_TaxID=238;  
 RN [1]  
 RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.  
 RC STRAIN-ELDER / ATCC 33958.  
 RX MEDLINE=95142655; PubMed=7840643;  
 RA Tarentino A.L., Quinones G., Hauser C.R., Changchlen L.M.,  
 RA Plummer T.H. Jr.,  
 RT "Molecular cloning and sequence analysis of Flavobacterium  
 meningosepticum glycosylasparaginase: a single gene encodes the alpha  
 and beta subunits."  
 RL Arch. Biochem. Biophys. 316:399-406(1995).  
 RN [2]  
 RP SEQUENCE OF 46-59 AND 197-211.  
 RX MEDLINE=94071939; PubMed=8250923;  
 RA Tarentino A.L., Plummer T.H. Jr.,  
 RT "The first demonstration of a procaryotic glycosylasparaginase."  
 RL Biochem. Biophys. Res. Commun. 197:179-186(1993).  
 RN [3]  
 RP X-RAY CRYSTALLOGRAPHY (2.32 ANGSTROMS).  
 RX MEDLINE=98200483; PubMed=9541410;  
 RA Xian J., Tarentino A.L., Grimwood B.G., Plummer T.H. Jr., Cul T.,  
 RA Guan C., van Roey P.,  
 RT "Crystal structure of glycosylasparaginase from Flavobacterium  
 meningosepticum."  
 RL Protein Sci. 7:774-781(1998).  
 CC -1- FUNCTION: CLEAVES THE GLCNAC-ASN BOND WHICH JOINS OLIGOSACCHARIDES  
 CC TO THE PEPTIDE OF ASPARAGINE-LINKED GLYCOPROTEINS. REQUIRES THAT  
 CC THE GLYCOSYLATED ASPARAGINE MOIETY IS NOT SUBSTITUTED ON  
 CC ITS AMINO (RI) AND CARBOXYL (R2) TERMINUS.  
 CC -1- CATALYTIC ACTIVITY: N4-(BETA-N-ACETYL-D-GLUCOSAMINYL)-L-ASPARAGINE  
 CC + H2O -> N-ACETYL-BETA-GLUCOSAMINYLAMINE + L-ASPARATE.  
 CC -1- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA CHAINS.  
 CC -1- SUBCELLULAR LOCATION: PERIPLASMIC.  
 CC -1- SIMILARITY: BELONGS TO THE ASPARAGINASE 2 FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC EMBL: U08028; AAA68868.1; -  
 DR PDB: 1AVY; 29-APR-98.  
 DR MEROPS: T02.001; -.

DR InterPro: IPR000246; -  
 DR Pfam: PF01112; Asparaginase\_2; 1.  
 DR Signal: Hydrolase; Periplasmic; 3D-structure.  
 FT SIGNAL 1 45  
 FT CHAIN 46 196 GLYCOSYLASPARAGINASE, ALPHA CHAIN.  
 FT CHAIN 197 340 GLYCOSYLASPARAGINASE, BETA CHAIN.  
 FT DISULFID 113 125  
 FT DISULFID 213 277  
 FT ACT\_SITE 197 197 OR 315.  
 FT ACT\_SITE 197 197 BY SIMILARITY.  
 SO SEQUENCE 340 AA; 37262 MW; 4C56E5061B4E53D7 CRC64;

Query Match 73.9%; Score 34; DB 1; Length 340;  
 Best Local Similarity 62.5%; Pred. No. 33;  
 Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 GGYC-VODG 8  
 DB 312 GAYCIODG 319

RESULT 6  
 ID STHA\_MYCTU STANDARD; PRT; 468 AA.  
 AC 007212;  
 DT 01-OCT-2000 (Rel. 40, Created)  
 DT 01-OCT-2000 (Rel. 40, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE PROBABLE SOLUBLE PYRIDINE NUCLEOTIDE TRANSHYDROGENASE (EC 1.6.1.1)  
 DE (STH) (NAD(P)(+)) TRANSHYDROGENASE [B-SPECIFIC].  
 GN STHA OR RV2713 OR MYC05A6.34.  
 OS Mycobacterium tuberculosis.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.  
 OX NCBI\_TaxID=1773;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-H37RV;  
 RX MEDLINE=98295987; PubMed=9634230;  
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
 RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. II, Tekela E.,  
 RA Baddock R., Basham D., Brown D., Chillingworth T., Connor R.,  
 RA Davies R., Devlin K., Fellwell T., Gentles S., Hamlin N., Holroyd S.,  
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,  
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
 RA Rutter S., Seeger K., Skelton S., Squares S., Squires R., Sulston J.E.,  
 RA Taylor K., Whitehead S., Barrett B.G.,  
 RT "Deciphering the biology of Mycobacterium tuberculosis from the  
 RT complete genome sequence."  
 RL Nature 393:537-544(1998).  
 CC -1- FUNCTION: CONVERSION OF NADPH, GENERATED BY PERIPHERAL CATABOLIC  
 CC PATHWAYS, TO NADH, WHICH CAN ENTER THE RESPIRATORY CHAIN FOR  
 CC ENERGY GENERATION (BY SIMILARITY).  
 CC -1- CATALYTIC ACTIVITY: NADPH + NAD(+) -> NADP(+) + NADH.  
 CC -1- COFACTOR: FAD (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE PYRIDINE NUCLEOTIDE-DISULFIDE  
 CC OXIDOREDUCTASES CLASS-I.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC EMBL: Z96072; CAB09473.1; -  
 DR HSP: P11959; IABD.  
 DR TubercuList: RV2713; -  
 DR InterPro: IPR000205; -  
 DR InterPro: IPR001100; -  
 DR Pfam: PF00070; pyr\_redox; 1.



DR PRINTS: PR00411; PNDRTASEI.  
 KW Oxidoreductase: Flavoprotein: FAD (ADP PART) (BY SIMILARITY).  
 FT NP\_BIND 5 35 FAD (ADP PART) (BY SIMILARITY).  
 FT NP\_BIND 300 310 FAD (FLAVIN PART) (BY SIMILARITY).  
 SO SEQUENCE 468 AA; 50754 MW; D9E737C41C2898CF CRC64;

Query Match 73.9%; Score 34; DB 1; Length 468;  
 Best Local Similarity 75.0%; Pred. No. 45;  
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 GGVGVQDG 8  
 |||||  
 DB 39 GGVGVNMG 46

RESULT 7  
 FP2\_MVTGA STANDARD: PRT: 473 AA.  
 ID FP2\_MVTGA  
 AC 025464;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE ADHESIVE PLAQUE MATRIX PROTEIN 2 PRECURSOR (FOOT PROTEIN 2) (MGFP2)  
 DE (MGFP-2).  
 GN FP2.  
 OS Mytilus galloprovincialis (Mediterranean mussel).  
 OC Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorpha; Mytiloida;  
 OC Mytilidae; Mytilidae; Mytilus.  
 OX NCBI\_TaxID=29158;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=FOOT;  
 RX MEDLINE=95204464; PubMed=7896812;  
 RA Inoue K., Takeuchi Y., Miki D., Odo S.;  
 RT "Mussel adhesive plaque protein gene is a novel member of epidermal growth factor-like gene family."  
 RL J. Biol. Chem. 270:6698-6701(1995).  
 CC -1- FUNCTION: PROVIDES ADHESIVENESS TO THE MUSSEL'S FOOT. MUSSELS PRODUCE ONE OF THE STRONGEST WATER INSOLUBLE GLUE. THE MUSSEL'S ADHESIVE IS A BUNDLE OF THREADS, CALLED A BYSSUS, FORMED BY A FIBROUS COLLAGENOUS CORE COATED WITH ADHESIVE PROTEINS.  
 CC -1- SUBCELLULAR LOCATION: SECRETED.  
 CC -1- TISSUE SPECIFICITY: PRODUCED BY THE BYSSAL GLAND.  
 CC -1- DEVELOPMENTAL STAGE: EXPRESSION STARTS AT THE PEDIVELIGER, FOOT FORMATION, STAGE.  
 CC -1- PTM: SOME TYROSINES ARE HYDROXYLATED (THUS PRODUCING DOPA - 3,4-DIHYDROXYPHENYLALANINE).  
 CC -1- SIMILARITY: CONTAINS 11 EGF-LIKE DOMAINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: D43794; BAA07852.1; -.  
 DR InterPro: IPR000561; -.  
 DR Pfam: PF00008; EGF\_11.  
 DR PROSITE: PS00010; ASX\_HYDROXYL. 2.  
 DR PROSITE: PS00022; EGF\_11.  
 DR PROSITE: PS01186; EGF\_2; 10.  
 DR GlycoProtein: EGF-like domain; Repeat; Signal; Hydroxylation.  
 FT SIGNAL 1 17  
 FT CHAIN 18 473 ADHESIVE PLAQUE MATRIX PROTEIN 2.  
 FT DOMAIN 45 81 EGF-LIKE 1.  
 FT DOMAIN 82 117 EGF-LIKE 2.  
 FT DOMAIN 118 134 EGF-LIKE 3.  
 FT DOMAIN 155 191 EGF-LIKE 4.  
 FT DOMAIN 192 228 EGF-LIKE 5.  
 FT DOMAIN 229 265 EGF-LIKE 6.

FT DOMAIN 266 301 EGF-LIKE 7.  
 FT DOMAIN 302 340 EGF-LIKE 8.  
 FT DOMAIN 342 378 EGF-LIKE 9.  
 FT DOMAIN 383 420 EGF-LIKE 10.  
 FT DOMAIN 425 461 EGF-LIKE 11.  
 FT MOD\_RES 23 23 DOPA.  
 FT MOD\_RES 31 31 DOPA.  
 FT MOD\_RES 36 36 DOPA.  
 FT MOD\_RES 43 43 DOPA.  
 FT CARBOHYD 93 93 N-LINKED (GLCNAc... ) (POTENTIAL).  
 FT DISULFID 49 60 BY SIMILARITY.  
 FT DISULFID 54 69 BY SIMILARITY.  
 FT DISULFID 71 80 BY SIMILARITY.  
 FT DISULFID 86 97 BY SIMILARITY.  
 FT DISULFID 91 106 BY SIMILARITY.  
 FT DISULFID 108 117 BY SIMILARITY.  
 FT DISULFID 122 133 BY SIMILARITY.  
 FT DISULFID 127 143 BY SIMILARITY.  
 FT DISULFID 145 154 BY SIMILARITY.  
 FT DISULFID 159 170 BY SIMILARITY.  
 FT DISULFID 164 180 BY SIMILARITY.  
 FT DISULFID 182 191 BY SIMILARITY.  
 FT DISULFID 196 207 BY SIMILARITY.  
 FT DISULFID 201 217 BY SIMILARITY.  
 FT DISULFID 219 228 BY SIMILARITY.  
 FT DISULFID 233 244 BY SIMILARITY.  
 FT DISULFID 238 254 BY SIMILARITY.  
 FT DISULFID 256 265 BY SIMILARITY.  
 FT DISULFID 270 281 BY SIMILARITY.  
 FT DISULFID 275 290 BY SIMILARITY.  
 FT DISULFID 292 301 BY SIMILARITY.  
 FT DISULFID 306 317 BY SIMILARITY.  
 FT DISULFID 311 328 BY SIMILARITY.  
 FT DISULFID 330 339 BY SIMILARITY.  
 FT DISULFID 346 357 BY SIMILARITY.  
 FT DISULFID 351 366 BY SIMILARITY.  
 FT DISULFID 368 377 BY SIMILARITY.  
 FT DISULFID 387 399 BY SIMILARITY.  
 FT DISULFID 393 408 BY SIMILARITY.  
 FT DISULFID 410 419 BY SIMILARITY.  
 FT DISULFID 429 440 BY SIMILARITY.  
 FT DISULFID 434 449 BY SIMILARITY.  
 FT DISULFID 451 460 BY SIMILARITY.  
 SO SEQUENCE 473 AA; 51772 MW; BA76B86C3BA49A0F CRC64;

Query Match 73.9%; Score 34; DB 1; Length 473;  
 Best Local Similarity 62.5%; Pred. No. 45;  
 Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 GGVGVQDG 8  
 ||:|  
 DB 241 GGVGVNMG 248

RESULT 8  
 CAH\_DUNSA STANDARD: PRT: 589 AA.  
 ID CAH\_DUNSA  
 AC P54212;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 01-OCT-1996 (Rel. 34, Last annotation update)  
 DE CARBONIC ANHYDRASE (EC 4.2.1.1) (CARBONATE DEHYDRATASE).  
 DE DCA.  
 GN Dunaliella salina.  
 OS Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;  
 OC Dunaliellaceae; Dunaliella.  
 OC NCBI\_TaxID=3046;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=96279304; PubMed=8663366;  
 RA Fisher M., Gokman I., Pick U., Zamir A.;  
 "A salt-resistant plasma membrane carbonic anhydrase is induced by

RT salt in Dunalbella salina.<sup>2</sup>  
 CC J. Biol. Chem. 271:17718-17723(1996).  
 CC -1- FUNCTION: REVERSIBLE HYDRATION OF CARBON DIOXIDE.  
 CC -1- CATALYTIC ACTIVITY: H(2)CO(3) -> CO(2) + H(2)O.  
 CC -1- INDUCTION: BY SALT.  
 CC -1- SIMILARITY: BELONGS TO THE EUKARYOTIC-TYPE CARBONIC ANHYDRASE  
 CC FAMILY.  
 CC -----  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC of the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: U53811; AAC49378.1; -  
 CC HSSP: P00918; ICRA  
 CC InterPro: IPR001148; -  
 CC Pfam: PF00194; carb\_anhydase; 2.  
 CC Lyase; Zinc.  
 CC DOMAIN 390 589 CATALYTIC.  
 CC FT METAL 418 418 ZINC (CATALYTIC) (BY SIMILARITY).  
 CC FT METAL 420 420 ZINC (CATALYTIC) (BY SIMILARITY).  
 CC FT METAL 440 440 ZINC (CATALYTIC) (BY SIMILARITY).  
 CC SEQUENCE 589 AA; 64257 MW; 0CC2A6F42C121171 CRC64;

Query Match 73.9%; Score 34; DB 1; Length 589;  
 Best Local Similarity 75.0%; Pred. No. 55;  
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GGVCYVDG 8  
 Db 72 GGVCVNTG 79

RESULT 9  
 UG33\_SQUTU STANDARD: PRT; 788 AA.  
 AC 043847;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE GLYCOGEN [STARCH] SYNTHASE PRECURSOR (EC 2.4.1.11) (GBSSI) (GRANULE-  
 DE BOUND STARCH SYNTHASE II) (FRAGMENT).  
 OS Solanum tuberosum (Potato).  
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 OC Magnoliophyta; eudicotyledons; core eudicots; Asteridae; eunasterids I;  
 OC Solanales; Solanaceae; Solanum.  
 NC NCB1\_TaxID=4113;  
 RN RA  
 RA Edwards A., Marshall J., Sidebottom C., Visser R.G.F., Smith A.M.,  
 RA Martin C.;  
 RA "Biochemical and molecular characterization of a novel starch  
 RT synthase from potato tubers."  
 RT Plant J. 8:283-294(1995).  
 RL  
 CC -1- FUNCTION: ACCOUNTS FOR ONLY 10 TO 15% OF THE TOTAL SOLUBLE STARCH  
 CC SYNTHASE ACTIVITY IN TUBERS.  
 CC -1- CATALYTIC ACTIVITY: UDP-GLUCOSE + (1,4-ALPHA-D-GLUCOSYL)[N] -  
 CC UDP + (1,4-ALPHA-D-GLUCOSYL)[N+1].  
 CC -1- PATHWAY: STARCH BIOSYNTHESIS.  
 CC -1- SUBCELLULAR LOCATION: CHLOROPLAST/AMYLOPLAST, SOLUBLE AND GRANULE-  
 CC BOUND.  
 CC -1- SIMILARITY: BELONGS TO THE BACTERIAL/PLANT GLYCOGEN SYNTHASE  
 CC FAMILY.  
 CC -----  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: X87968; CAA61241.1; -  
 CC KM Glycogen biosynthesis: Transferease; Glycosyltransferase;  
 CC Transit peptide; Chloroplast; Amyloplast; Starch biosynthesis.  
 CC FT NON\_TER 1 1  
 CC FT TRANSIT 1 65 CHLOROPLAST.  
 CC FT CHAIN 66 788 GLYCOGEN [STARCH] SYNTHASE.  
 CC FT BINDING 310 310 UDP-GLUCOSE (BY SIMILARITY).  
 CC FT VARIANT 71 71 S -> D.  
 CC SEQUENCE 788 AA; 87890 MW; 8DBB90611E862B7B CRC64;

Query Match 73.9%; Score 34; DB 1; Length 788;  
 Best Local Similarity 75.0%; Pred. No. 71;  
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GGVCYVDG 8  
 Db 419 GGVCYVDG 426

RESULT 10  
 BG1L\_ASPAC STANDARD: PRT; 860 AA.  
 ID BG1L\_ASPAC  
 AC P48825;  
 DT 01-FEB-1996 (Rel. 33, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE BETA-D-GLUCOSIDASE 1 PRECURSOR (EC 3.2.1.21) (GENTIOBIASE) (CELLULOBIASE)  
 DE (BETA-D-GLUCOSIDE GLUCOHYDROLASE).  
 OS Aspergillus aculeatus.  
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
 OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.  
 NC NCB1\_TaxID=5053;  
 RN RA  
 RA Ooi T., Arai M.;  
 RA "Cloning and sequencing of the cDNA encoding beta-glucosidase 1 from  
 RT Aspergillus aculeatus."  
 RT Gene 173:287-288(1996).  
 RL  
 CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF TERMINAL, NON-REDUCING BETA-D-  
 CC GLUCOSE RESIDUES WITH RELEASE OF BETA-D-GLUCOSE.  
 CC -1- PATHWAY: CELLULOSE DEGRADATION.  
 CC -1- SIMILARITY: BELONGS TO FAMILY 3 OF GLYCOSYL HYDROLASES.  
 CC -----  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: D64088; BAA10968.1; -  
 CC InterPro: IPR001764; -  
 CC InterPro: IPR002772; -  
 CC Pfam: PF00933; Glyco\_hydro\_3; 1.  
 CC Pfam: PF00195; Glyco\_hydro\_3.C; 1.  
 CC PRINTS: PR00133; GLHYDRLASE3.  
 CC PROSITE: PS00775; GLYCOSYL\_HYDROL\_F3; 1.  
 CC Hydrolyase; Glycosidase; Cellulose degradation; glycoprotein; signal.  
 CC FT SIGNAL 1 19  
 CC FT CHAIN 20 860 BETA-GLUCOSIDASE 1.  
 CC FT ACT\_SITE 280 280 BY SIMILARITY.  
 CC FT CARBOHYD 61 61 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 CC FT CARBOHYD 211 211 N-LINKED (GLCNAC. . .) (POTENTIAL).

```

FT CARBOHYD 252 252 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 315 315 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 322 322 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 354 354 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 387 387 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 442 442 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 523 523 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 542 542 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 564 564 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 658 658 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 668 668 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 690 690 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 712 712 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 860 AA; 93052 MW; 4B484778B0DFC694 CRC64;

Query Match
Best Local Similarity 73.9%; Score 34; DB 1; Length 860;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 GGCVCQD 7
   111111
Db 86 GGCICLD 92

RESULT 11
EF2_BLAHO STANDARD: PRT; 867 AA.
AC Q17152;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE ELONGATION FACTOR 2 (EF-2).
OS Blastocystis hominis.
OC Eukaryota; stramenopiles; Blastocystis.
OX NCBI_Taxid=12968;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HEB7-1;
RA Nakamura Y., Hashimoto T., Yoshikawa H., Kamatsht T., Nakamura F.,
RA Okamoto K.I., Hasegawa M.;
RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: THIS PROTEIN PROMOTES THE GTP-DEPENDENT TRANSLOCATION
CC OF THE NASCENT PROTEIN CHAIN FROM THE A-SITE TO THE P-SITE OF THE
CC RIBOSOME.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -1- PTM: PHOSPHORYLATION BY EF-2 KINASE COMPLETELY INACTIVATES EF-2
CC (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE GTP-BINDING ELONGATION FACTOR FAMILY.
CC EF-G/EF-2 SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D79219; BAA11469.1; -.
DR HSSP; P07157; IAIP.
DR InterPro; IPR000640; -.
DR InterPro; IPR000795; -.
DR Pfam; PF00679; EFG_C; 1.
DR Pfam; PF00009; GTP_EFTU; 1.
DR PROSITE; PS00301; EFATOR_GTP; 1.
KM Elongation factor: GTP-binding; Protein biosynthesis; Phosphorylation.
FT NP_BIND 26 33 GTP (BY SIMILARITY).
FT NP_BIND 120 124 GTP (BY SIMILARITY).
FT NP_BIND 176 179 GTP (BY SIMILARITY).
FT MOD_RES 57 57 PHOSPHORYLATION (BY SIMILARITY).
FT MOD_RES 59 59 PHOSPHORYLATION (BY SIMILARITY).
FT MOD_RES 723 723 DIPTAMIDE (BY SIMILARITY).

```

```

SQ SEQUENCE 867 AA; 96349 MW; EE4BBE57DEC61A0F CRC64;

Query Match
Best Local Similarity 73.9%; Score 34; DB 1; Length 867;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GGCVCQ 6
   111111
Db 149 GGCVCQ 154

RESULT 12
DDR1_RAT STANDARD: PRT; 910 AA.
ID DDR1_RAT
AC Q63474;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE EPITHELIAL DISCOIDIN DOMAIN RECEPTOR 1 PRECURSOR (EC 2.7.1.112)
DE (TYROSINE-PROTEIN KINASE CAK) (CELL ADHESION KINASE) (TYROSINE KINASE
DE DDR) (DISCOIDIN RECEPTOR TYROSINE KINASE) (PROTEIN-TYROSINE KINASE
DE PTK-3).
GN DDR1 OR EDDR1 OR PTK3.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_Taxid=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=Brain;
RA MEDLINE-94173920; Pubmed-8127887;
RA Sanchez M.P., Tapley P., Saini S.S., He B., Pulido D., Barbaclad M.;
RT "Multiple tyrosine protein kinases in rat hippocampal neurons:
RT isolation of Ptk-3, a receptor expressed in proliferative zones of
RT the developing brain.";
RL Proc. Natl. Acad. Sci. U.S.A. 91:1819-1823(1994).
CC -1- FUNCTION: MAY BE INVOLVED IN CELL-CELL INTERACTIONS AND
CC RECOGNITION (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE - ADP +
CC PROTEIN TYROSINE PHOSPHATE.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -1- TISSUE SPECIFICITY: VARIOUS EMBRYONIC AND ADULT TISSUES; ALSO
CC PROLIFERATIVE ZONES OF THE DEVELOPING BRAIN; HIPPOCAMPAL NEURONS.
CC -1- SIMILARITY: WITH THE CONSERVED CATALYTIC DOMAINS OF TYR-PROTEIN
CC KINASES.
CC -1- SIMILARITY: CONTAINS 1 P5/8 TYPE C DOMAIN.
CC -1- SIMILARITY: BELONGS TO THE INSULIN RECEPTOR FAMILY OF TYROSINE-
CC PROTEIN KINASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; I26525; AAA21089.1; -.
DR HSSP; P11362; IFIGI.
DR InterPro; IPR000421; -.
DR InterPro; IPR000719; -.
DR InterPro; IPR001245; -.
DR InterPro; IPR002011; -.
DR Pfam; PF00754; P5_F8_type_C; 1.
DR Pfam; PF00069; Pkinase; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE; PS00239; RECEPTOR_TYR_KIN_II; 1.
DR PROSITE; PS01285; FAS8C_1; 1.
DR PROSITE; PS01286; FAS8C_2; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KM Transferase: Tyrosine-protein kinase; Glycoprotein; Signal;
Phosphorylation; Transmembrane; Receptor; ATP-binding.

```

FT SIGNAL 1 19 POTENTIAL.  
 FT CHAIN 20 910 EPITHELIAL DISCOIDIN DOMAIN RECEPTOR 1.  
 FT DOMAIN 20 413 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 414 440 POTENTIAL.  
 FT DOMAIN 414 910 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 32 186 F5/8 TYPE C (PHOSPHOLIPID-BINDING, POTENTIAL).  
 FT DOMAIN 378 412 GLY/PRO-RICH.  
 FT DOMAIN 473 598 GLY/PRO-RICH.  
 FT DOMAIN 607 902 PROTEIN KINASE.  
 FT NP\_BIND 613 621 ATP (BY SIMILARITY).  
 FT BINDING 652 652 ATP (BY SIMILARITY).  
 FT ACN\_SITE 763 763 BY SIMILARITY.  
 FT DISULFID 32 186 BY SIMILARITY.  
 FT MOD\_RRS 510 510 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
 FT MOD\_RRS 789 789 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
 FT MOD\_RRS 793 793 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
 FT MOD\_RRS 794 794 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
 FT CARBOHYD 212 212 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 261 261 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 371 371 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 391 391 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT SEQUENCE 910 AA; 101164 MW; 7E7FFAIDCB029806 CRC64;

Query Match 73.9%; Score 34; DB 1; Length 910;  
 Best Local Similarity 100.0%; Pred. No. 81;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GVCVQD 7  
 Db 686 GVCVQD 691

RESULT 13  
 DDRL\_MOUSE  
 ID DDRL\_MOUSE STANDARD; PRT: 911 AA.  
 AC Q03146;  
 DT 01-OCT-1994 (Rel. 30, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE EPITHELIAL DISCOIDIN DOMAIN RECEPTOR 1 PRECURSOR (EC 2.7.1.112)  
 DE (TYROSINE-PROTEIN KINASE) (CELL ADHESION KINASE) (TYROSINE KINASE  
 DE DDR) (DISCOIDIN RECEPTOR TYROSINE KINASE) (PROTEIN-TYROSINE KINASE  
 DE MPK-6).  
 GN DDR1 OR EDDR1 OR CAK OR MPK6.  
 OS Mus musculus (mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RP [1]  
 RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.  
 RC STRAIN=C57BL;  
 RC MEDLINE=96204002; PubMed=8622863;  
 RA Peter J.L., Jing S.Q., Wong T.W.;  
 RT Identification of two isoforms of the Cak receptor kinase that are  
 RT coexpressed in breast tumor cell lines.";  
 RL Oncogene 12:1469-1477(1996).  
 RN [2]  
 RP SEQUENCE OF 766-822 FROM N.A.  
 RC STRAIN=C57BL; TISSUE=Embryonic brain;  
 RX MEDLINE=93096484; PubMed=1281307;  
 RA Giaraldi-Hebenstreit P., Nieto M.A., Frain M., Mattei M.-G.,  
 RA Chestier A., Wilkinson B.G., Charney P.;  
 RT "An Eph-related receptor protein tyrosine kinase gene segmentally  
 RT expressed in the developing mouse hindbrain.";  
 RL Oncogene 7:2499-2506(1992).  
 CC -1- FUNCTION: MAY BE INVOLVED IN CELL-CELL INTERACTIONS AND  
 CC RECOGNITION (BY SIMILARITY).  
 CC -1- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE - ADP +  
 CC PROTEIN TYROSINE PHOSPHATE.  
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.  
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; CAK I (SHOWN HERE) AND CAK II;

CC ARE PRODUCED BY ALTERNATIVE SPLICING. THEY ONLY DIFFER BY THE  
 CC ABSENCE OF A 37 RESIDUES SEGMENT.  
 CC -1- TISSUE SPECIFICITY: THE PREDOMINANT ISOFORM CAK I, IS EXPRESSED IN  
 CC DEVELOPING EMBRYO AND ADULT BRAIN; CAK II IS EXPRESSED IN VARIOUS  
 CC EPITHELIAL CELLS.  
 CC -1- SIMILARITY: WITH THE CONSERVED CATALYTIC DOMAINS OF TYR-PROTEIN  
 CC KINASES.  
 CC -1- SIMILARITY: CONTAINS 1 F5/8 TYPE C DOMAIN.  
 CC -1- SIMILARITY: BELONGS TO THE INSULIN RECEPTOR FAMILY OF TYROSINE-  
 CC PROTEIN KINASES.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: L57509; AAB05209.1; -;  
 CC EMBL: X57240; CAA40516.1; -;  
 CC PIR: S30502; S30502.  
 CC HSP: P11362; IFEI.  
 CC MGD: WGI:99216; Cak.  
 CC InterPro: IPR000421; -;  
 CC InterPro: IPR000719; -;  
 CC InterPro: IPR001245; -;  
 CC InterPro: IPR002011; -;  
 CC Pfam: PF00754; F5\_F8-type-C; 1.  
 CC Pfam: PF00069; Pkinase; 1.  
 CC PROSITE: PS00107; PROTEIN\_KINASE\_ATP; FALSE\_NEG.  
 CC PROSITE: PS00109; PROTEIN\_KINASE\_TYR; 1.  
 CC PROSITE: PS00111; PROTEIN\_KINASE\_DOM; 1.  
 CC PROSITE: PS00239; RECEPTOR\_TYR\_KIN\_II; 1.  
 CC PROSITE: PS01285; PA58C\_1; 1.  
 CC PROSITE: PS01286; PA58C\_2; 1.  
 CC Transferase; Tyrosine-protein kinase; Glycoprotein; Signal;  
 CC Phosphorylation; Transmembrane; Receptor; ATP-binding;  
 CC Alternative splicing.  
 KW SIGNAL 1 19 POTENTIAL.  
 KW CHAIN 20 911 EPITHELIAL DISCOIDIN DOMAIN RECEPTOR 1.  
 KW DOMAIN 20 414 EXTRACELLULAR (POTENTIAL).  
 KW TRANSMEM 415 441 POTENTIAL.  
 KW DOMAIN 442 911 CYTOPLASMIC (POTENTIAL).  
 KW DOMAIN 32 186 F5/8 TYPE C (PHOSPHOLIPID-BINDING, POTENTIAL).  
 KW DOMAIN 378 412 GLY/PRO-RICH.  
 KW DOMAIN 474 598 GLY/PRO-RICH.  
 KW DOMAIN 608 903 PROTEIN KINASE.  
 KW NP\_BIND 613 622 ATP (BY SIMILARITY).  
 KW BINDING 653 653 ATP (BY SIMILARITY).  
 KW ACN\_SITE 764 764 BY SIMILARITY.  
 KW DISULFID 32 186 BY SIMILARITY.  
 KW MOD\_RRS 511 511 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
 KW MOD\_RRS 790 790 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
 KW MOD\_RRS 794 794 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
 KW MOD\_RRS 795 795 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
 KW CARBOHYD 213 213 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 KW CARBOHYD 262 262 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 KW CARBOHYD 372 372 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 KW CARBOHYD 392 392 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 KW VARSPIC 503 539 MISSING (IN ISOFORM CAK II).  
 SQ SEQUENCE 911 AA; 101160 MW; DBB7FE03DD79510 CRC64;

Query Match 73.9%; Score 34; DB 1; Length 911;  
 Best Local Similarity 100.0%; Pred. No. 81;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GVCVQD 7  
 Db 687 GVCVQD 692

RESULT 14  
 DDRL\_HUMAN STANDARD; PRT; 913 AA.  
 AC 008345; Q14196; Q16562;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE EPITHELIAL DISCOIDIN DOMAIN RECEPTOR 1 PRECURSOR (EC 2.7.1.112)  
 DE (TYROSINE-PROTEIN KINASE CAK) (CELL ADHESION KINASE) (TYROSINE KINASE  
 DE DDR) (DISCOIDIN RECEPTOR TYROSINE KINASE) (TRK E) (PROTEIN-TYROSINE  
 DE KINASE RTK 6).  
 GN DDR1 OR EDDR1 OR CAK OR TRKE OR RTK6.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-Petal liver;  
 RX MEDLINE-9413417; PubMed-8302582;  
 RA Perez J.L., Shen X., Finkernagel S., Scliorra L., Jenkins N.A.,  
 RA Gilbert D.J., Copeland N.G., Wong T.W.;  
 RT "Identification and chromosomal mapping of a receptor tyrosine kinase  
 RT with a putative phospholipid binding sequence in its ectodomain.";  
 RL Oncogene 9:211-219(1994).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-Placenta;  
 RX MEDLINE-93296201; PubMed-8390675;  
 RA Johnson J.D., Edman J.C., Rutter W.J.;  
 RT "A receptor tyrosine kinase found in breast carcinoma cells has an  
 RT extracellular discoidin I-like domain";  
 RL Proc. Natl. Acad. Sci. U.S.A. 90:5677-5681(1993).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-97131588; PubMed-8977099;  
 RA Sakuma S., Tada M., Saye H., Sawamura Y., Shinohe Y., Abe H.;  
 RT "Receptor protein tyrosine kinase DDR is up-regulated by p53  
 RT protein.";  
 RL FEBS Lett. 398:165-169(1996).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-96389017; PubMed-8796349;  
 RA Playford M.P., Butler R.J., Wang X.C., Katso R.M., Cooke I.E.,  
 RA Ganesan T.S.;  
 RT "The genomic structure of discoidin receptor tyrosine kinase.";  
 RL Genome Res. 6:620-627(1996).  
 RN [5]  
 RP SEQUENCE FROM N.A. (SHORT FORM).  
 RC TISSUE-Lung;  
 RX MEDLINE-96204002; PubMed-8622863;  
 RA Perez J.L., Jing S.O., Wong T.W.;  
 RT "Identification of two isoforms of the Cak receptor kinase that are  
 RT coexpressed in breast tumor cell lines";  
 RL Oncogene 12:1469-1477(1996).  
 RN [6]  
 RP SEQUENCE FROM N.A. (SHORT FORM).  
 RC TISSUE-Brain, and Keratinocytes;  
 RX MEDLINE-94043265; PubMed-8226977;  
 RA di Marco E., Cutuli N., Guerra L., Cancedda R., de Luca M.;  
 RT "Molecular cloning of trke, a novel trk-related putative tyrosine  
 RT kinase receptor isolated from normal human keratinocytes and widely  
 RT expressed by normal human tissues";  
 RL J. Biol. Chem. 268:24295-24295(1993).  
 RN [7]  
 RP SEQUENCE FROM N.A. (SHORT FORM).  
 RC TISSUE-Ovary;  
 RX MEDLINE-95151638; PubMed-7848919;  
 RA Laval S., Butler R., Shelling A.N., Hanby A.M., Poulson R.,  
 RA Ganesan T.S.;  
 RT "Isolation and characterization of an epithelial-specific receptor  
 RT tyrosine kinase from an ovarian cancer cell line.";

RL Cell Growth Differ. 5:1173-1183(1994).  
 CC -1- FUNCTION: MAY BE INVOLVED IN CELL-CELL INTERACTIONS AND  
 CC RECOGNITION.  
 CC -1- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE - ADP +  
 CC PROTEIN TYROSINE PHOSPHATE.  
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.  
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS, CAK I (SHOWN HERE) AND CAK II;  
 CC ARE PRODUCED BY ALTERNATIVE SPLICING. THEY ONLY DIFFER BY THE  
 CC ABSENCE OF A 37 RESIDUES SEGMENT.  
 CC -1- TISSUE SPECIFICITY: EXPRESSED AT LOW LEVELS IN MOST ADULT TISSUES  
 CC AND IS HIGHEST IN THE BRAIN AND LUNG. ABUNDANT IN BREAST CARCINOMA  
 CC CELL LINES.  
 CC -1- DOMAIN: THE GLY/PRO-RICH DOMAINS MAY BE REQUIRED FOR AN UNUSUAL  
 CC GEOMETRY OF INTERACTION WITH LIGAND OR SUBSTRATES.  
 CC -1- SIMILARITY: TO OTHER PROTEIN-TYROSINE KINASES IN THE CATALYTIC  
 CC DOMAIN.  
 CC -1- SIMILARITY: CONTAINS 1 F5/8 TYPE C DOMAIN.  
 CC -1- SIMILARITY: BELONGS TO THE INSULIN RECEPTOR FAMILY OF TYROSINE-  
 CC PROTEIN KINASES.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>  
 CC or send an email to [license@isb-sdb.ch](mailto:license@isb-sdb.ch)).  
 CC -----  
 DR EMBL: L11315; AAA02866.1; -  
 DR EMBL: L20817; AAA18019.1; -  
 DR EMBL: U48705; AAC50917.1; -  
 DR EMBL: X88208; CAA66871.1; -  
 DR EMBL: X99023; CAA66871.1; JOINED.  
 DR EMBL: X99024; CAA66871.1; JOINED.  
 DR EMBL: X99025; CAA66871.1; JOINED.  
 DR EMBL: X99026; CAA66871.1; JOINED.  
 DR EMBL: X99027; CAA66871.1; JOINED.  
 DR EMBL: X99028; CAA66871.1; JOINED.  
 DR EMBL: X99029; CAA66871.1; JOINED.  
 DR EMBL: X99030; CAA66871.1; JOINED.  
 DR EMBL: X99031; CAA66871.1; JOINED.  
 DR EMBL: X99032; CAA66871.1; JOINED.  
 DR EMBL: X99033; CAA66871.1; JOINED.  
 DR EMBL: X99034; CAA66871.1; JOINED.  
 DR EMBL: L57508; AAB05208.1; -  
 DR EMBL: X74979; CAA52915.1; -  
 DR EMBL: Z29093; CAA82335.1; -  
 DR HSSP: P11362; 1FGI.  
 DR MIM: 600408; -  
 DR InterPro: IPR000421; -  
 DR InterPro: IPR000719; -  
 DR InterPro: IPR001245; -  
 DR InterPro: IPR002011; -  
 DR Pfam: PF00754; F5\_F8\_type\_C; 1.  
 DR Pfam: PF00069; pkinase; 1.  
 DR ProSite: PS00107; PROTEIN\_KINASE\_ATP; FALSE\_NEG.  
 DR ProSite: PS00109; PROTEIN\_KINASE\_TYR; 1.  
 DR ProSite: PS00111; PROTEIN\_KINASE\_DOM; 1.  
 DR ProSite: PS00239; RECEPTOR\_TYR\_KIN\_II; 1.  
 DR ProSite: PS01285; FA58C\_1; 1.  
 DR ProSite: PS01286; FA58C\_2; 1.  
 KW Transferase; Tyrosine-protein kinase; Glycoprotein; Signal;  
 KW Phosphorylation; Receptor; ATP-binding;  
 KW Alternative splicing.  
 FT SIGNAL 1 18  
 FT CHAIN 19 913  
 FT DOMAIN 19 416  
 FT TRANSMEM 417 443  
 FT DOMAIN 444 913  
 FT DOMAIN 31 185  
 FT DOMAIN 377 415  
 FT DOMAIN 476 601  
 FT POTENTIAL.  
 FT EPITHELIAL DISCOIDIN DOMAIN RECEPTOR 1.  
 FT EXTRACELLULAR (POTENTIAL).  
 FT POTENTIAL.  
 FT CYTOPLASMIC (POTENTIAL).  
 FT F5/8 TYPE C (PHOSPHOLIPID-BINDING,  
 FT POTENTIAL).  
 FT GLY/PRO-RICH.  
 FT GLY/PRO-RICH.

FT DOMAIN 610 905 PROTEIN KINASE.  
 FT N\_BIND 616 924 ATP (BY SIMILARITY).  
 FT BINDING 655 655 ATP (BY SIMILARITY).  
 FT ACT\_SITE 766 766 BY SIMILARITY.  
 FT DISULFID 31 184 BY SIMILARITY.  
 FT MOD\_RES 513 513 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
 FT MOD\_RES 792 792 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
 FT MOD\_RES 796 796 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
 FT MOD\_RES 797 797 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
 FT CARBOHYD 211 211 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 260 266 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 371 371 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 394 394 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT VARSPLIC 306 342 MISSING (IN ISOFORM CAK II).  
 FT CONFLICT 94 94 L -> V (IN REF. 2 AND 3).  
 FT CONFLICT 833 833 L -> V (IN REF. 2 AND 3).  
 FT CONFLICT 847 867 OLIDOVLENCEFEPRDGRQ -> SAHRRAHREGRGVLP  
 FT  
 SO SEQUENCE 913 AA, 101127 MW: G96913EA906C481E CRC64;  
 GGGRA (IN REF. 4).  
 GGGRA (IN REF. 4).

Query Match 73.9%; Score 34; DB 1; Length 913;  
 Best Local Similarity 100.0%; Pred. No. 82;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 2 GVCYOD 7  
 DB 689 GVCYOD 694

RESULT 15  
 CALL CHICK STANDARD; PRT: 1453 AA.  
 AC P02457;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 01-OCT-1989 (Rel. 12, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE COLLAGEN ALPHA 1(I) CHAIN PRECURSOR.  
 GN COL1A1.  
 OS Gallus gallus (Chicken).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 CC Gallus.  
 OK NCBI\_TaxID=9031;  
 RN [1]  
 RP SEQUENCE OF 1-153 FROM N.A.  
 RX MEDLINE=88056316; PubMed=3678834;  
 RA Fliner M.H., Boedtker H., Doty P.;  
 RT "Construction and characterization of cDNA clones encoding the 5' end  
 of the chicken pro alpha 1(I) collagen mRNA.";  
 RL Gene 56:71-78(1987).  
 RN [2]  
 RP SEQUENCE OF 1-144 FROM N.A.  
 RX MEDLINE=88007542; PubMed=2820966;  
 RA Fliner M.H., Aho S., Gerstenfeld L.C., Boedtker H., Doty P.;  
 RT "Unusual DNA sequences located within the promoter region and the  
 first intron of the chicken pro-alpha 1(I) collagen gene.";  
 RL J. Biol. Chem. 262:13323-13332(1987).  
 RN [3]  
 RP SEQUENCE OF 152-1187.  
 RX MEDLINE=8231995; PubMed=7093229;  
 RA Hlgberger J.H., Corbett C., Dixit S.N., Yu W., Seyer J.M.,  
 RA Kang A.H., Gross J.;  
 RT "Amino acid sequence of chick skin collagen alpha 1(I)-C8B and the  
 complete primary structure of the helical portion of the chick skin  
 collagen alpha 1(I) chain.";  
 RL Biochemistry 21:2048-2055(1982).  
 RN [4]  
 RP SEQUENCE OF 1200-1205.  
 RX MEDLINE=72243016; PubMed=5047697;  
 RA Byre D.R., Glimcher M.J.;  
 RT "Evidence for a previously undetected sequence at the carboxyterminus  
 of the alpha 1 chain of chicken bone collagen.";

RL Biochem. Biophys. Res. Commun. 48:720-726(1972).  
 RN [5]  
 RP SEQUENCE OF 981-1453 FROM N.A.  
 RX MEDLINE=81160715; PubMed=6927845;  
 RA Fuller F., Boedtker H.;  
 RT "Sequence determination and analysis of the 3' region of chicken pro-  
 alpha 1(I) and pro-alpha 2(I) collagen messenger ribonucleic acids  
 including the carboxy-terminal propeptide sequences.";  
 RL Biochemistry 20:996-1006(1981).  
 RN [6]  
 RP SEQUENCE OF 1311-1453 FROM N.A.  
 RX MEDLINE=80134546; PubMed=6987088;  
 RA Showalter A.M., Pesciotta D.M., Elkenberry E.F., Yamamoto T.,  
 RA Pastan I., Decrombrughe B., Fietzek P.P., Olsen B.R.;  
 RT "Nucleotide sequence of a collagen cDNA-fragment coding for the  
 carboxyl end of pro alpha 1(I)-chains.";  
 RL FEBS Lett. 111:61-65(1980).  
 CC -1- FUNCTION: TYPE I COLLAGEN IS A MEMBER OF GROUP I COLLAGEN  
 (FIBRILLAR FORMING COLLAGEN).  
 CC -1- SUBUNIT: TRIMERS OF ONE ALPHA 2(I) AND TWO ALPHA 1(I) CHAINS.  
 CC -1- TISSUE SPECIFICITY: FORMS THE FIBRILS OF TENDON, LIGAMENTS AND  
 BONES. IN BONES THE FIBRILS ARE MINERALIZED WITH CALCIUM  
 HYDROXYAPATITE.  
 CC -1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING  
 UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.  
 CC -1- SIMILARITY: CONTAINS 1 WFC DOMAIN.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC  
 DR EMBL: M17839; AAA48704.1; -  
 DR EMBL: M17838; AAA48704.1; JOINED.  
 DR EMBL: V00401; CAA23695.1; -  
 DR EMBL: M10571; AAA48671.1; ALT\_SEQ.  
 DR EMBL: M17607; AAA48672.1; -  
 DR PIR: A02857; CGCHS.  
 DR PIR: A27179; A27179.  
 DR PIR: A29367; A29367.  
 DR InterPro: IPR000087; -  
 DR InterPro: IPR000885; -  
 DR InterPro: IPR001007; -  
 DR Pfam: PF01410; COLF1; 1.  
 DR Pfam: PF01391; COLLAGEN; 18.  
 DR Pfam: PF00093; WFC; 1.  
 DR PROSITE: PS01208; WFC; 1.  
 KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;  
 KW Glycoprotein; Collagen; Signal.  
 FT SIGNAL 1 22  
 FT PROPEP 23 151 AMINO-TERMINAL PROPEPTIDE.  
 FT CHAIN 152 1205 COLLAGEN ALPHA 1(I) CHAIN.  
 FT PROPEP 1206 1453 C-TERMINAL PROPEPTIDE.  
 FT DOMAIN 31 89 WFC.  
 FT MOD\_RES 152 152 PYROLIDONE CARBOXYLIC ACID.  
 FT MOD\_RES 254 254 HYDROXYLATION (POTENTIAL).  
 FT MOD\_RES 851 851 HYDROXYLATION (POTENTIAL).  
 FT MOD\_RES 1081 1081 HYDROXYLATION (POTENTIAL).  
 FT MOD\_RES 1097 1097 HYDROXYLATION (POTENTIAL).  
 FT MOD\_RES 1153 1153 HYDROXYLATION (ONLY 3-HYDROXYPRO AND THE  
 ONLY HYDROXYLATED PRO IN POSITION X (IN  
 THE G-X-Y UNIT IN THE ALPHA 1(I) CHAIN)).  
 FT CONFLICT 1187 1187 F -> L (IN REF. 5).  
 FT CONFLICT 1441 1441 O -> H (IN REF. 6).  
 SO SEQUENCE 1453 AA, 137789 MW: 3BC6152134271F4D CRC64;

Query Match 73.9%; Score 34; DB 1; Length 1453;  
 Best Local Similarity 85.7%; Pred. No. 1,2e+02;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy	2	GVCVODG	8
Db	31	GSCVODG	37

Search completed: June 13, 2001, 14:21:46  
Job time: 806 sec

*This Page Blank (uspto)*



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 13, 2001, 14:20:33 ; Search time 225.85 Seconds  
(without alignments)  
4.152 Million cell updates/sec

Title: PCT-US01-05825A-14

Perfect score: 46

Sequence: 1 GGCVCQDG 8

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 374700 seqs, 117207915 residues

Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL.15:\*

- 1: sp\_archaea:\*
- 2: sp\_bacteria:\*
- 3: sp\_fungi:\*
- 4: sp\_human:\*
- 5: sp\_invertebrate:\*
- 6: sp\_mammal:\*
- 7: sp\_mhc:\*
- 8: sp\_organelle:\*
- 9: sp\_phage:\*
- 10: sp\_plant:\*
- 11: sp\_protist:\*
- 12: sp\_unclassified:\*
- 13: sp\_vertebrate:\*
- 14: sp\_virus:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	41	89.1	196	2	084353
2	41	89.1	196	2	09PK45
3	37	80.4	354	11	090U51
4	37	80.4	373	11	090YV1
5	37	80.4	380	4	013086
6	36	78.3	138	13	09PSF8
7	36	78.3	172	13	002529
8	36	78.3	217	4	014964
9	36	78.3	336	14	068405
10	36	78.3	633	10	09LZ06
11	36	78.3	680	14	011308
12	36	78.3	680	14	098187
13	36	78.3	713	13	090330
14	36	78.3	822	13	091288
15	36	78.3	922	13	090413
16	36	78.3	984	9	080102
17	36	78.3	1213	5	09VE06
18	36	78.3	1219	5	09YIA8
19	35	76.1	323	2	09L8F5

20	35	76.1	399	2	09R3V6	09r3v6 vibrio chol
21	35	76.1	399	2	09L706	09l706 vibrio chol
22	35	76.1	402	1	026284	026284 methanobact
23	35	76.1	1031	10	065500	065500 arabidopsis
24	34	73.9	50	4	09UD35	09ud35 homo sapien
25	34	73.9	57	4	09UDL2	09udl2 homo sapien
26	34	73.9	103	4	099227	099227 homo sapien
27	34	73.9	119	6	077753	077753 canis famli
28	34	73.9	256	5	022208	022208 caenorhabdi
29	34	73.9	266	2	09PCV9	09pcv9 xyella fas
30	34	73.9	461	10	049610	049610 arabidopsis
31	34	73.9	468	2	007212	007212 mycobacteri
32	34	73.9	470	2	060154	060154 pelobacter
33	34	73.9	537	2	005459	005459 mycobacteri
34	34	73.9	732	10	048859	048859 zea mays (m
35	34	73.9	751	10	09SPH5	09sph5 manihot esc
36	34	73.9	753	10	09LWY1	09lwy1 arabidopsis
37	34	73.9	792	10	09MAC8	09mac8 arabidopsis
38	34	73.9	793	3	000903	000903 gaeumannomy
39	34	73.9	861	11	09QW58	09qw58 mus sp. mot
40	34	73.9	911	11	035407	035407 mus musculu
41	34	73.9	1487	4	014047	014047 homo sapien
42	34	73.9	1491	13	091718	091718 xenopus lae
43	34	73.9	1548	11	062040	062040 mus musculu
44	34	73.9	1722	5	019350	019350 caenorhabdi
45	34	73.9	2327	13	091B67	091b67 xenopus lae

#### ALIGNMENTS

RESULT	ID	084353	PRELIMINARY:	PRT:	196 AA.
AC	084353				
DT	01-NOV-1998	(TREMBLrel. 08, Created)			
DT	01-NOV-1998	(TREMBLrel. 08, Last sequence update)			
DT	01-MAY-2000	(TREMBLrel. 13, Last annotation update)			
DE	MAF-TYPE PROTEIN.				
GN	MAF.				
OS	Chlamydia trachomatis.				
OC	Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.				
OX	NCBI_Taxid=813;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=D/UM-3/CX;				
RX	MEDLINE=99000809; PubMed=9784136;				
RA	Stephens R.S., Kaiman S., Lammell C.J., Fan J., Marathe R., Aravind L., Mitchell W.P., Olinger L., Tatunov R.L., Zhao Q., Koonin E.V., Davis R.W.;				
RT	"Genome sequence of an obligate intracellular pathogen of humans: Chlamydia trachomatis."				
RL	Science 282:754-759(1998).				
DR	EMBL: AE001308; AAC67944.1; "				
SD	SEQUENCE 196 AA; 21978 MW; EDB6CFD52F93073 CRC64;				
Query Match		89.1%;	Score 41;	DB 2;	Length 196;
Best Local Similarity		87.5%;	Pred. No. 1.9;		
Matches	7;	Conservative	0;	Mismatches	1;
				Indels	0;
				Gaps	0;
QY	1	GGCVCQDG 8			
DB	150	GGCVCQDG 157			
RESULT	2				
ID	09PK45	PRELIMINARY:	PRT:	196 AA.	
AC	09PK45				
DT	01-OCT-2000	(TREMBLrel. 15, Created)			
DT	01-OCT-2000	(TREMBLrel. 15, Last sequence update)			
DT	01-OCT-2000	(TREMBLrel. 15, Last annotation update)			

DE MAF PROTEIN.  
 GN TC0628.  
 OS Chlamydia muridarum.  
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
 NC NCBI\_TaxID=83560;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MOPN / NIGG;  
 RX MEDLINE=20150255; Pubmed=10684935;  
 RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,  
 White O., Hickey E.K., Peterson J., Uterback T., Berry K.,  
 Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,  
 Gilm M., Nelson W., Deboy R., Kolonay J., McClarty G., Salzberg S.L.,  
 Eisen J., Fraser C.M.;  
 RA Genome sequences of Chlamydia trachomatis MOPN and Chlamydia  
 pneumoniae AR39.";  
 RT Pneumoniae AR39.";  
 RL Nucleic Acids Res. 28:1397-1406(2000).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Read T.D., Brunham R., Shen C., Gill S.R., Heidelberg J.F., White O.,  
 Hickey E.K., Peterson J., Umayam L.A., Uterback T., Berry K.,  
 Bass S., Linher K., Weidman J., Khouri H., Craven B., Bowman C.,  
 Dodson R., Gilm M., Nelson W., Deboy R., Kolonay J., McClarty G.,  
 Salzberg S.L., Eisen J., Fraser C.M.;  
 RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AEO02331; AAF39457.1; -.  
 DR TIGR: TC0628; -.  
 SQ SEQUENCE 196 AA; 21730 MW; 9B86B059E52D78A5 CRC64;

Query Match 89.1%; Score 41; DB 2; Length 196;  
 Best Local Similarity 87.5%; Pred. No. 1.9;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGVCVODG 8  
 11111111  
 Db 150 GGVCVODG 157

RESULT 3  
 ID 09JUS1 PRELIMINARY; PRT; 354 AA.  
 AC 09JUS1;  
 DT 01-OCT-2000 (TREMblrel. 15, Created)  
 DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)  
 DT 01-OCT-2000 (TREMblrel. 15, Last annotation update)  
 DE TOMOREGULIN-1 PRECURSOR (FRAGMENT).  
 GN TR-1.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NC NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Eib D.W., Holling T.M., Zwilgen A., Dewulf N., de Groot E.,  
 van den Eijnden-van Raaij A.J.M., Huybrecock D., Martens G.J.;  
 RT \*Expression of the follistatin/EGF-containing transmembrane protein  
 M7365 (tomoregulin-1) during mouse development.";  
 RT Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AJ400622; CAB90827.1; -.  
 KW Signal.  
 FT NON\_TER 1  
 FT SIGNAL <1 11 POTENTIAL.  
 FT CHAIN 12 354 TOMOREGULIN-1.  
 SQ SEQUENCE 354 AA; 38202 MW; A10F53A328B666 CRC64;

Query Match 80.4%; Score 37; DB 11; Length 354;  
 Best Local Similarity 75.0%; Pred. No. 20;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 GGVCVODG 8  
 11111111

Db 58 GGVCCKEDG 65

RESULT 4  
 ID 09QYV1 PRELIMINARY; PRT; 373 AA.  
 AC 09QYV1;  
 DT 01-MAY-2000 (TREMblrel. 13, Created)  
 DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)  
 DT 01-OCT-2000 (TREMblrel. 13, Last annotation update)  
 DE NCI PROTEIN.  
 GN NCI.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 NC NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=BRAIN;  
 RA Kugler S., Baehr M.;  
 RT \*Reverse transcription of a highly G+C rich mRNA 5'end by Tth  
 polymerase resolved inversions and deletions which were generated by  
 RT MMLV reverse transcriptase.";  
 RT Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AJ250730; CAB60131.1; -.  
 DR HSSP: P00998; ITGS.  
 DR INTERPRO: IPR000561; -.  
 DR INTERPRO: IPR002350; -.  
 DR PFAM: PF00050; kazal; 2.  
 DR PROSITE: PS00022; EGF\_1; UNKNOWN\_1.  
 DR PROSITE: PS01186; EGF\_2; 1.  
 SQ SEQUENCE 373 AA; 40143 MW; BDB8BC681B10280E CRC64;

Query Match 80.4%; Score 37; DB 11; Length 373;  
 Best Local Similarity 75.0%; Pred. No. 21;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGVCVODG 8  
 11111111  
 Db 78 GGVCCKEDG 85

RESULT 5  
 ID 013086 PRELIMINARY; PRT; 380 AA.  
 AC 013086;  
 DT 01-NOV-1996 (TREMblrel. 01, Created)  
 DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)  
 DT 01-OCT-2000 (TREMblrel. 15, Last annotation update)  
 DE TRANSMEMBRANE PROTEIN PRECURSOR.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 NC NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=BRAIN;  
 RA Eib D.W., Martens G.J.;  
 RT Submitted (APR-1995) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: U19678; AAA64622.1; -.  
 DR HSSP: P00998; ITGS.  
 DR INTERPRO: IPR000561; -.  
 DR INTERPRO: IPR001239; -.  
 DR INTERPRO: IPR002350; -.  
 DR PFAM: PF00050; kazal; 2.  
 DR PRINTS: PR00290; KAZALINHTR.  
 DR PROSITE: PS00022; EGF\_1; UNKNOWN\_1.  
 DR PROSITE: PS01186; EGF\_2; 1.  
 KW Signal; Transmembrane.  
 FT SIGNAL 1 39 POTENTIAL.  
 FT CHAIN 40 380 TRANSMEMBRANE PROTEIN.  
 FT TRANSMEM 331 351 POTENTIAL.

SO SEQUENCE 380 AA; 40885 MW; E220F55683B0A368 CRC64;

Query Match 80.4%; Score 37; DB 4; Length 380;  
Best Local Similarity 75.0%; Pred. No. 22;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 GVCVODG 8  
|||:|  
DB 85 GVCVKEDG 92

RESULT 6

09PSF8 PRELIMINARY; PRT; 138 AA.

AC 09PSF8;  
DT 01-MAY-2000 (TREMBLrel. 13, Created)  
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)  
DE FIBROBLAST GROWTH FACTOR RECEPTOR CPE-FGFR.

OS Gallus gallus (Chicken).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Actinopterygii; Atherinomorpha; Phasianidae; Phasianinae;

OC Gallus.

OC NCBI\_TaxID=9031;

OX NCBI\_TaxID=9031;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=94346318; PubMed=8067293;

RA Fujiwara M.;

RT "[Analysis of fibroblast growth factor receptor genes expressed in the  
retinal pigment epithelium of chick embryos by reverse transcription

RT polymerase chain reaction].";

RL J. Jpn. Ophthalmol. Soc. 98:625-629(1994).

DR HSSP; P12931; 1FMK.

DR INTERPRO; IPR000719; -.

DR INTERPRO; IPR001245; -.

DR PFAM; PF00069; pkinase.1.

DR PRINTS; PR00109; TYRKINASE.

DR PROSITE; PSS0011; PROTEIN\_KINASE\_DOM; 1.

SO SEQUENCE 138 AA; 16109 MW; 74465170F678AE58 CRC64;

Query Match 78.3%; Score 36; DB 13; Length 138;  
Best Local Similarity 85.7%; Pred. No. 12;

Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 GVCVODG 8  
|||:|  
DB 19 GVCVODG 25

RESULT 7

002529 PRELIMINARY; PRT; 172 AA.

AC 002529;

DT 01-NOV-1996 (TREMBLrel. 01, Created)

DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)

DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)

DE FIBROBLAST GROWTH FACTOR RECEPTOR (EC 2.7.1.112) (FGF-R) (FRAGMENT).

GN MF4.

OS Oryzias latipes (Medaka fish).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;

OC Acanthomorphi; Acanthopterygii; Percormorpha; Atherinomorpha;

OC Belontiiformes; Adrianchthyidae; Oryziinae; Oryzias.

OC NCBI\_TaxID=8090;

OX NCBI\_TaxID=8090;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=93093167; PubMed=1459248;

RA Emori Y., Yasuoka A., Saigo K.;

RT "Identification of four FGF receptor genes in Medaka fish (Oryzias  
latipes).";

RL FEBS Lett. 314:176-178(1992).

CC -1- FUNCTION: RECEPTOR FOR ACIDIC AND BASIC FIBROBLAST GROWTH FACTORS.  
CC -1- SIMILARITY: BELONGS TO THE FIBROBLAST GROWTH FACTOR RECEPTOR  
FAMILY.

DR EMBL; D13553; BAA02752.1; -.

DR PIR; S27022; S27022.

DR HSSP; P12931; 1FMK.

DR INTERPRO; IPR000719; -.

DR INTERPRO; IPR001245; -.

DR PFAM; PF00069; pkinase.1.

DR PRINTS; PR00109; TYRKINASE.

DR PROSITE; PSS00109; PROTEIN\_KINASE\_TYR; 1.

DR PROSITE; PSS0011; PROTEIN\_KINASE\_DOM; 1.

KW Receptor; Tyrosine-protein kinase; ATP-binding; Transferase;

KW Phosphorylation; Growth factor binding.

FT NON\_TER 1

FT ACT\_SITE 132 132

FT BINDING 23 23

FT MOD\_RES 163 163

FT NON\_TER 172 172

SO SEQUENCE 172 AA; 19404 MW; DEA21F383673D655 CRC64;

Query Match 78.3%; Score 36; DB 13; Length 172;  
Best Local Similarity 85.7%; Pred. No. 15;

Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 GVCVODG 8  
|||:|  
DB 58 GVCVODG 64

RESULT 8

014964 PRELIMINARY; PRT; 217 AA.

AC 014964;

DT 01-NOV-1996 (TREMBLrel. 01, Created)

DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)

DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)

DE RAB-RELATED GTP-BINDING PROTEIN (FRAGMENT).

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

OC NCBI\_TaxID=9606;

OX NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=97237046; PubMed=9119394;

RA Stankovic T., Byrd P.J., Cooper P.R., McConville C.M., Munroe D.J.,

RA Riley J.H., Watts G.D.J., Ambrose H., McGuire G., Smith A.D.,

RA Sutcliffe A., Mills T., Taylor A.M.R.;

RT "Construction of a transcription map around the gene for ataxia  
telangiectasia: Identification of at least four novel genes.";

RL EMBL; X99962; CAA68227.1; -.

DR HSSP; P05713; 3RAB.

DR INTERPRO; IPR001806; -.

DR PFAM; PF00071; Ras; 1.

DR PRINTS; PR00449; RASTRNSFRMG.

FT NON\_TER 1

FT NON\_TER 217

SO SEQUENCE 217 AA; 24869 MW; AA9A29731F42B8F5 CRC64;

Query Match 78.3%; Score 36; DB 4; Length 217;  
Best Local Similarity 62.5%; Pred. No. 19;

Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 GVCVODG 8  
|||:|  
DB 182 GVCVODG 189

RESULT 9

068405

ID 068405 PRELIMINARY; PRT; 336 AA.  
 AC 068405;  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
 DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)  
 DE ORF UL151.  
 OS Human cytomegalovirus.  
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
 OC Betaherpesvirinae; Cytomegalovirus.  
 OX NCBI\_TaxID=10359;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=TOLEDO;  
 RX MEDLINE=96099416; PubMed=8523595;  
 RA Cha T.A., Tom E., Kemble G.W., Duke G.M., Mocarski E.S., Spaete R.R.;  
 RT "Human cytomegalovirus clinical isolates carry at least 19 genes not  
 found in laboratory strains.";  
 RL J Virol 70:78-83(1996).  
 DR EMBL: U3331; AAA8582.1; -;  
 SQ SEQUENCE 336 AA; 35116 MW; 9F865E5019F69D0C CRC64;

Query Match 78.3%; Score 36; DB 14; Length 336;  
 Best Local Similarity 75.0%; Pred. NO. 30;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 GGCVCVODG 8  
 1111:11  
 Db 72 GGCVCEDG 79

RESULT 10  
 ID 09L206 PRELIMINARY; PRT; 633 AA.  
 AC 09L206;  
 DT 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)  
 DE HYPOTHETICAL 71.7 KDA PROTEIN.  
 GN F16L2\_50.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;  
 OC Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Jordan N., Bangert S., Wiedemann R., Voss H., Unsel M., Mewes H.W.,  
 RA Rudd S., Lemcke K., Mayr K.F.X., Quetier F., Salanoubat M.;  
 RL Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA EU Arabidopsis sequencing project;  
 RL Submitted (APR-2000) to the EMBL/Genbank/DBJ databases.  
 DR EMBL: AL162459; CAB82808.1; -;  
 KW Hypothetical protein  
 SQ SEQUENCE 633 AA; 71695 MW; 2758B9C38381DF14 CRC64;

Query Match 78.3%; Score 36; DB 10; Length 633;  
 Best Local Similarity 75.0%; Pred. NO. 56;  
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 GGCVCVODG 8  
 1111:11  
 Db 624 GGCVCVODG 631  
 RESULT 11  
 ID 011308 PRELIMINARY; PRT; 640 AA.  
 AC 011308;  
 DT 01-JUL-1997 (TREMBLrel. 04, Created)

DT 01-JUL-1997 (TREMBLrel. 04, Last sequence update)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)  
 DE SIMILAR TO VARIOUS C15L AND VACCINIA FILI.  
 GN B-W,N.L.8.  
 OS Molluscum contagiosum virus subtype 1 (MCV1).  
 OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;  
 OC Molluscipoxvirus.  
 OX NCBI\_TaxID=10280;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97093414; PubMed=8938976;  
 RA Martin-Gallardo A., Moratilla M., Funes J.M., Agromayor M., Nunez A.,  
 RA Varas A.J., Collado M., Valencia A., Lopez-Esteban J.L.,  
 RA Esteban M.;  
 RT "Sequence analysis of a Molluscum contagiosum virus DNA region which  
 includes the gene encoding protein kinase 2 and other genes with  
 unique organization.";  
 RL Virus Genes 13:19-29(1996).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Moratilla M., Agromayor M., Nunez A., Funes J.M., Varas A.J.,  
 RA Lopez-Esteban J.L., Esteban M., Martin-Gallardo A.;  
 RL Virus Genes 0:0-0(0).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA Moratilla M., Agromayor M., Nunez A., Funes J.M., Varas A.J.,  
 RA Lopez-Esteban J.L., Collado M., Payon M., Martin-Gallardo A.;  
 RL Submitted (JAN-1997) to the EMBL/Genbank/DBJ databases.  
 DR EMBL: U86894; AAB57938.1; -;  
 DR INTERPRO: IPR002221; -;  
 DR PROSITE: PS00317; 4\_DISULFIDE\_CORE; UNKNOWN\_1.  
 SQ SEQUENCE 640 AA; 68996 MW; 23752F0B08B49BEA CRC64;

Query Match 78.3%; Score 36; DB 14; Length 640;  
 Best Local Similarity 85.7%; Pred. NO. 57;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 GGCVCVODG 8  
 111111  
 Db 151 GACVCVODG 157

RESULT 12  
 ID 098187 PRELIMINARY; PRT; 680 AA.  
 AC 098187;  
 DT 01-FEB-1997 (TREMBLrel. 02, Created)  
 DT 01-FEB-1997 (TREMBLrel. 02, Last sequence update)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)  
 DE MC018L.  
 GN MC018L.  
 OS Molluscum contagiosum virus subtype 1 (MCV1).  
 OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;  
 OC Molluscipoxvirus.  
 OX NCBI\_TaxID=10280;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=96325459; PubMed=8670425;  
 RA Senkevich T.G., Bugert J.J., Sisler J.R., Koonin E.V., Darai G.,  
 RA Moss B.;  
 RT "Genome sequence of a human tumorigenic poxvirus: prediction of  
 specific host response-evasion genes.";  
 RL Science 273:813-816(1996).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Senkevich T.G., Bugert J.J., Sisler J.R., Koonin E.V., Darai G.,  
 RA Moss B.;  
 RL Submitted (JUN-1996) to the EMBL/Genbank/DBJ databases.  
 DR EMBL: U60315; AAC55146.1; -;  
 DR INTERPRO: IPR002221; -;  
 DR PROSITE: PS00317; 4\_DISULFIDE\_CORE; UNKNOWN\_1.  
 SQ SEQUENCE 680 AA; 72398 MW; 12CB3BD9E61AA72B CRC64;

Query Match 78.3%; Score 36; DB 14; Length 680;  
 Best Local Similarity 85.7%; Pred. No. 60;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 GVCVODG 8  
 DB 151 GACVODG 157

RESULT 13

ID 090330 PRELIMINARY; PRT: 713 AA.

AC 090330;  
 DT 01-NOV-1996 (TREMBlrel. 01, Created)  
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)  
 DE FIBROBLAST GROWTH FACTOR RECEPTOR PRECURSOR.

CN FREK.  
 OS Coturnix coturnix japonica (Japanese quail).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 OC Coturnix.  
 OC NCBI\_TaxID=93934;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=EMBRYO;  
 RA MEDLINE=94215505; PubMed=8162862;  
 RT "Distinct developmental expression of a new avian fibroblast growth factor receptor."  
 RT Development 120:683-694(1994).  
 RL EMBL: X76885; CAA54213.1; -.  
 DR HSSP: P12931; 1FMK.  
 DR INTERPRO: IPR000719; -.  
 DR INTERPRO: IPR001245; -.  
 DR INTERPRO: IPR003006; -.  
 DR PFAM: PF00047; 1g; 2.  
 DR PFAM: PF00069; pkinase; 1.  
 DR PRINTS: PR00107; TYRKINASE.  
 DR PROSITE: PS00107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE: PS00109; PROTEIN\_KINASE\_TYR; 1.  
 DR PROSITE: PS50011; PROTEIN\_KINASE\_DOM; 1.  
 KW Signal.  
 FT SIGNAL 1 19  
 FT CHAIN 20 713  
 FT SEQUENCE 713 AA; 80169 MW; B7DE1BF422A85F66 CRC64;  
 SO FIBROBLAST GROWTH FACTOR RECEPTOR.

Query Match 78.3%; Score 36; DB 13; Length 713;  
 Best Local Similarity 85.7%; Pred. No. 63;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 GVCVODG 8  
 DB 450 GVCVODG 456

RESULT 14  
 ID 091288 PRELIMINARY; PRT: 822 AA.

AC 091288;  
 DT 01-NOV-1996 (TREMBlrel. 01, Created)  
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
 DT 01-MAY-2000 (TREMBlrel. 13, Last annotation update)  
 DE FIBROBLAST GROWTH FACTOR RECEPTOR.  
 OS Pleurodeles waltl (Iberian ribbed newt).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Amphibia; Batrachia; Caudata; Salamandroidae; Salamandridae;  
 OC Pleurodeles.  
 OC NCBI\_TaxID=8319;  
 RN [1]

RP SEQUENCE FROM N.A.  
 RX MEDLINE=93130775; PubMed=1483392;  
 RA Shi D.L., Felge J.J., Riou J.F., Desjone D.W., Boucaut J.C.;

RT "Differential expression and regulation of two distinct fibroblast growth factor receptors during early development of the urodele amphibian Pleurodeles waltl."  
 RT amphibian Pleurodeles waltl.  
 RT Development 116:261-273(1992).  
 RL EMBL: X65059; CAA6192.1; -.

DR HSSP: P06239; 3LCK.  
 DR INTERPRO: IPR000719; -.  
 DR INTERPRO: IPR001245; -.  
 DR INTERPRO: IPR003006; -.  
 DR PFAM: PF00047; 1g; 3.  
 DR PFAM: PF00069; pkinase; 1.  
 DR PRINTS: PR00107; TYRKINASE.  
 DR PROSITE: PS00107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE: PS00109; PROTEIN\_KINASE\_TYR; 1.  
 DR PROSITE: PS50011; PROTEIN\_KINASE\_DOM; 1.  
 SO SEQUENCE 822 AA; 92068 MW; 3EC4BA4BDC9AB81A CRC64;

Query Match 78.3%; Score 36; DB 13; Length 822;  
 Best Local Similarity 85.7%; Pred. No. 73;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 GVCVODG 8  
 DB 557 GVCVODG 563

RESULT 15

ID 090413 PRELIMINARY; PRT: 922 AA.

AC 090413;  
 DT 01-NOV-1996 (TREMBlrel. 01, Created)  
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)  
 DE FIBROBLAST GROWTH FACTOR RECEPTOR 4.  
 GN FGFR4.  
 OS Brachydanio rerio (zebrafish) (Zebra danio).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;  
 OC Cypriniformes; Cyprinidae; Rasbora; Danio.  
 OC NCBI\_TaxID=7955;  
 RN [1]

RP SEQUENCE FROM N.A.  
 RX MEDLINE=96172998; PubMed=8589434;  
 RA Thisse B., Thisse C., Weston J.A.;  
 RT "Novel FGF receptor (Z-FGFR4) is dynamically expressed in mesoderm and neurectoderm during early zebrafish embryogenesis."  
 RT Dev. Dyn. 203:377-391(1995).  
 RL EMBL: U23839; AAA96816.1; -.  
 DR HSSP: P08631; 1AD5.  
 DR ZFIN: ZDB-GENE-980526-488; fgfr4.  
 DR INTERPRO: IPR000719; -.  
 DR INTERPRO: IPR001245; -.  
 DR INTERPRO: IPR003006; -.  
 DR PFAM: PF00047; 1g; 4.  
 DR PFAM: PF00069; pkinase; 1.  
 DR PRINTS: PR00107; TYRKINASE.  
 DR PROSITE: PS00107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE: PS00109; PROTEIN\_KINASE\_TYR; 1.  
 DR PROSITE: PS50011; PROTEIN\_KINASE\_DOM; 1.  
 SO SEQUENCE 922 AA; 103464 MW; 9560EFBEDFEDCF5 CRC64;

Query Match 78.3%; Score 36; DB 13; Length 922;  
 Best Local Similarity 85.7%; Pred. No. 82;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 GVCVODG 8  
 DB 649 GVCVODG 655

Wed Jun 13 14:59:46 2001

pct-us01-05825a-14.rspt

Page 6

Search completed: June 13, 2001, 14:20:34  
Job time: 734 sec

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 13, 2001, 14:16:35 ; Search time 118.55 Seconds  
(without alignments)  
1.296 Million cell updates/sec

Title: PCT-US01-05825A-14

Perfect score: 46  
Sequence: 1 GGCYQDC 8

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 185757 seqs, 19210857 residues

Total number of hits satisfying chosen parameters: 185757

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA:\*

- 1: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep:\*
- 2: /cgn2\_6/ptodata/2/1aa/5B.COMB.pep:\*
- 3: /cgn2\_6/ptodata/2/1aa/6A.COMB.pep:\*
- 4: /cgn2\_6/ptodata/2/1aa/6B.COMB.pep:\*
- 5: /cgn2\_6/ptodata/2/1aa/PTUS.COMB.pep:\*
- 6: /cgn2\_6/ptodata/2/1aa/Backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	37	80.4	380	3	US-08-468-846-2
2	36	78.3	115	2	US-07-903-029-4
3	36	78.3	115	2	US-07-903-029-5
4	36	78.3	336	1	US-08-414-926A-26
5	36	78.3	336	2	US-08-926-922-26
6	36	78.3	336	3	US-09-253-682-26
7	35	76.1	400	2	US-08-624-601-8
8	34	73.9	50	1	US-08-336-343A-17
9	34	73.9	317	2	US-08-336-343A-18
10	34	73.9	558	4	US-08-701-191A-25
11	34	73.9	669	4	US-08-836-567-6
12	34	73.9	669	4	US-08-941-445A-9
13	34	73.9	767	4	US-08-836-567-8
14	34	73.9	913	3	US-08-445-640-4
15	34	73.9	913	3	US-08-170-558-4
16	34	73.9	913	3	US-08-447-314-4
17	34	73.9	913	3	US-08-445-461-4
18	34	73.9	919	1	US-08-336-343A-2
19	33	71.7	512	3	US-09-015-296-3
20	33	71.7	1153	1	US-08-097-997A-14
21	33	71.7	1153	4	US-08-665-574C-14
22	33	71.7	1153	4	US-08-946-994-14
23	32	69.6	201	1	US-08-471-570-2
24	32	69.6	292	2	US-08-701-191A-40
25	32	69.6	299	2	US-08-701-191A-13
26	32	69.6	300	2	US-08-701-191A-31
27	32	69.6	309	2	US-08-701-191A-9

28	32	69.6	310	2	US-08-701-191A-1	Sequence 1, Appl
29	32	69.6	310	2	US-08-701-191A-6	Sequence 6, Appl
30	32	69.6	310	2	US-08-701-191A-7	Sequence 7, Appl
31	32	69.6	313	1	US-08-278-089A-17	Sequence 17, Appl
32	32	69.6	313	2	US-08-838-957A-16	Sequence 16, Appl
33	32	69.6	315	2	US-08-701-191A-2	Sequence 2, Appl
34	32	69.6	351	4	US-07-791-931-8	Sequence 8, Appl
35	32	69.6	351	4	US-08-701-191A-3	Sequence 3, Appl
36	32	69.6	378	1	US-08-070-165F-8	Sequence 8, Appl
37	32	69.6	378	2	US-08-885-418-8	Sequence 8, Appl
38	32	69.6	388	1	US-08-070-165F-4	Sequence 4, Appl
39	32	69.6	388	2	US-08-885-418-4	Sequence 4, Appl
40	32	69.6	526	1	US-08-446-038B-17	Sequence 17, Appl
41	32	69.6	581	1	US-08-446-010B-17	Sequence 17, Appl
42	32	69.6	581	1	US-08-805-445-17	Sequence 17, Appl
43	32	69.6	581	2	US-08-064-067D-17	Sequence 17, Appl
44	32	69.6	581	2	US-09-066-208-17	Sequence 17, Appl
45	32	69.6	581	2	US-09-066-208-17	Sequence 17, Appl

ALIGNMENTS

RESULT 1  
US-08-468-846-2  
; Sequence 2, Application US/08468846  
; Patent No. 6074839  
; GENERAL INFORMATION:  
; APPLICANT: Meissner, Paul  
; APPLICANT: Fuldner, Rebecca  
; APPLICANT: Fel-Wel, Ying  
; APPLICANT: Adams, Mark  
; TITLE OF INVENTION: TRANSFORMING GROWTH FACTOR ALPHA HI  
; NUMBER OF SEQUENCES: 15  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN, CECCHI,  
; ADDRESSEE: STUART & OLSTEIN  
; STREET: 6 Becker Farm Road  
; CITY: Roseland  
; STATE: New Jersey  
; COUNTRY: USA  
; ZIP: 07068  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/468,846  
; FILING DATE: 06-JUN-1995  
; CLASSIFICATION: 514  
; PRIORITY APPLICATION DATA:  
; APPLICATION NUMBER: US 08/208,008  
; FILING DATE: 08-MAR-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ferraro, Gregory D.  
; REGISTRATION NUMBER: 36,134  
; REFERENCE/DOCKET NUMBER: 325800-465  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 201-994-1700  
; TELEFAX: 201-994-1744  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 380 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-468-846-2

Query Match 80.4%; Score 37; DB 3; Length 380;  
Best local Similarity 75.0%; Pred. No. 38;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGVGVODG 8  
|||||  
Db 85 GGVCKEDG 92

## RESULT 2

US-07-903-029-4  
; Sequence 4, Application US/07903029  
; Patent No. 5969097  
; GENERAL INFORMATION:  
; APPLICANT: Wiegand, Roger C.  
; APPLICANT: Currie, Mark C.  
; TITLE OF INVENTION: Human Guanylin  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Dennis A. Bennett, Monsanto Co., A3SG  
; STREET: 800 N. Lindbergh Blvd.  
; CITY: St. Louis  
; STATE: Missouri  
; COUNTRY: USA  
; ZIP: 63167  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentln Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/903,029  
; FILING DATE: 19920623  
; CLASSIFICATION: 530  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Bennett, Dennis A.  
; REGISTRATION NUMBER: 34,547  
; REFERENCE/DOCKET NUMBER: 07-21(872)A  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (314)694-5402  
; TELEFAX: (314)694-9009  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 115 amino acids  
; TYPE: AMINO ACID  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-07-903-029-4

Query Match 78.38; Score 36; DB 2; Length 115;  
Best Local Similarity 87.58; Pred. No. 17;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGVGVODG 8  
|||||  
Db 20 GGVTVODG 27

## RESULT 3

US-07-903-029-5  
; Sequence 5, Application US/07903029  
; Patent No. 5969097  
; GENERAL INFORMATION:  
; APPLICANT: Wiegand, Roger C.  
; APPLICANT: Currie, Mark C.  
; TITLE OF INVENTION: Human Guanylin  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Dennis A. Bennett, Monsanto Co., A3SG  
; STREET: 800 N. Lindbergh Blvd.  
; CITY: St. Louis  
; STATE: Missouri  
; COUNTRY: USA

ZIP: 63167  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentln Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/903,029  
; FILING DATE: 19920623  
; CLASSIFICATION: 530  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Bennett, Dennis A.  
; REGISTRATION NUMBER: 34,547  
; REFERENCE/DOCKET NUMBER: 07-21(872)A  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (314)694-5402  
; TELEFAX: (314)694-9009  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 115 amino acids  
; TYPE: AMINO ACID  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-07-903-029-5

Query Match 78.38; Score 36; DB 2; Length 115;  
Best Local Similarity 87.58; Pred. No. 17;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGVGVODG 8  
|||||  
Db 20 GGVTVODG 27

RESULT 4  
US-08-414-926A-26  
; Sequence 26, Application US/08414926A  
; Patent No. 5721354  
; GENERAL INFORMATION:  
; APPLICANT: Spaete, Richard  
; APPLICANT: Cha, Tai-An  
; TITLE OF INVENTION: NOVEL HUMAN CYTOMEGALOVIRUS  
; NUMBER OF SEQUENCES: 27  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Cooley Godward Castro Huddleson & Tatum  
; STREET: 5 Palo Alto Square  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94306-2155  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentln Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/414,926A  
; FILING DATE: March 31, 1995  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Casert, Luann  
; REGISTRATION NUMBER: 31,822  
; REFERENCE/DOCKET NUMBER: AVIR-011/00US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-494-7622  
; TELEFAX: 415-857-0663  
; INFORMATION FOR SEQ ID NO: 26:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 336 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein



IMMEDIATE SOURCE:  
CLONE: tol.22  
FEATURE:  
NAME/KEY: Protein  
LOCATION: 1..336  
OTHER INFORMATION: /label= UL151  
US-08-414-926A-26

Query Match 78.3%; Score 36; DB 1; Length 336;  
Best Local Similarity 75.0%; Pred. No. 50;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 GVCVQDG 8  
1111:11  
DB 72 GVCVGEDG 79

RESULT 5  
US-08-926-922-26  
Sequence 26, Application US/08926922  
Patent No. 5925751  
GENERAL INFORMATION:  
APPLICANT: Spaete, Richard  
APPLICANT: Cha, Tai-An  
TITLE OF INVENTION: NOVEL HUMAN CYTOMEGALOVIRUS  
NUMBER OF SEQUENCES: 27  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Luann Cseer Attorney at Law  
STREET: 750 Arimo Avenue  
CITY: Oakland  
STATE: CA  
COUNTRY: USA  
ZIP: 94610  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/926,922  
FILING DATE: September 10, 1997  
CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:  
NAME: Cseer, Luann  
REGISTRATION NUMBER: 31,822  
REFERENCE/DOCKET NUMBER: AVIR 11A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 510-834-1448  
TELEFAX: 510-839-7810  
INFORMATION FOR SEQ ID NO: 26:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 336 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
IMMEDIATE SOURCE:  
CLONE: tol.22  
FEATURE:  
NAME/KEY: Protein  
LOCATION: 1..336  
OTHER INFORMATION: /label= UL151  
US-08-926-922-26

Query Match 78.3%; Score 36; DB 2; Length 336;  
Best Local Similarity 75.0%; Pred. No. 50;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 GVCVQDG 8  
1111:11  
DB 72 GVCVGEDG 79

RESULT 6  
US-09-253-682-26

Sequence 26, Application US/09253682  
Patent No. 6040170

GENERAL INFORMATION:

APPLICANT: Spaete, Richard

APPLICANT: Cha, Tai-An

TITLE OF INVENTION: NOVEL HUMAN CYTOMEGALOVIRUS

NUMBER OF SEQUENCES: 27

CORRESPONDENCE ADDRESS:

ADDRESSEE: Luann Cseer Attorney at Law

STREET: 750 Arimo Avenue

CITY: Oakland

STATE: CA

COUNTRY: USA

ZIP: 94610

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/253,682

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/926,922

FILING DATE: September 10, 1997

ATTORNEY/AGENT INFORMATION:

NAME: Cseer, Luann

REGISTRATION NUMBER: 31,822

REFERENCE/DOCKET NUMBER: AVIR 11A

TELECOMMUNICATION INFORMATION:

TELEPHONE: 510-834-1448

TELEFAX: 510-839-7810

INFORMATION FOR SEQ ID NO: 26:

SEQUENCE CHARACTERISTICS:

LENGTH: 336 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

IMMEDIATE SOURCE:

CLONE: tol.22

FEATURE:

NAME/KEY: Protein

LOCATION: 1..336

OTHER INFORMATION: /label= UL151  
US-09-253-682-26

Query Match 78.3%; Score 36; DB 3; Length 336;  
Best Local Similarity 75.0%; Pred. No. 50;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 GVCVQDG 8  
1111:11  
DB 72 GVCVGEDG 79

RESULT 7  
US-08-624-601-8

Sequence 8, Application US/08624601

Patent No. 5882653

GENERAL INFORMATION:

APPLICANT: Kaper Dr., James B.

APPLICANT: Levine Dr., Myron M.

TITLE OF INVENTION: Vibrio cholerae O1 (CVD11) and non-O1

(CVD112 and CVD112RM) serogroup vaccine strains, methods

of making same and products thereof

NUMBER OF SEQUENCES: 16

CORRESPONDENCE ADDRESS:

ADDRESSEE: Spencer & Frank

STREET: 1100 New York Ave. N.W. Suite 300 East  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20005  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/624,601  
FILING DATE: 08-Apr-1996  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Schellier Dr., John W.  
REGISTRATION NUMBER: 26,031  
REFERENCE/DOCKET NUMBER: BANC20019P2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)414-4000  
TELEFAX: (202)414-4040  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 400 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Vibrio cholerae  
STRAIN: El Tor 7946  
IMMEDIATE SOURCE:  
CLONE: zot  
US-08-624-601-8

Query Match 76.1%; Score 35; DB 2; Length 400;  
Best Local Similarity 75.0%; Pred. No. 86;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGCYVDG 8  
1:|||||  
DB 291 GRLCYVDG 298

RESULT 8  
US-08-336-343A-17  
Sequence 17, Application US/08336343A  
Patent No. 5677144  
GENERAL INFORMATION:  
APPLICANT: Ullrich, Axel  
APPLICANT: Alves, Frauke  
TITLE OF INVENTION: CCK-2, A No. 5677144el Receptor Tyrosine Kinase  
NUMBER OF SEQUENCES: 43  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pennie & Edmonds  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036-2711  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/336,343A  
FILING DATE: 08-NOV-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:

NAME: Coruzzi, Laura A.  
REGISTRATION NUMBER: 30,742  
REFERENCE/DOCKET NUMBER: 7683-065  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-9741/8864  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 17:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 44 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
US-08-336-343A-17

Query Match 73.9%; Score 34; DB 1; Length 44;  
Best Local Similarity 100.0%; Pred. No. 14;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GVCYVD 7  
1|||||  
DB 39 GVCYVD 44

RESULT 9  
US-08-336-343A-18  
Sequence 18, Application US/08336343A  
Patent No. 5677144  
GENERAL INFORMATION:  
APPLICANT: Ullrich, Axel  
APPLICANT: Alves, Frauke  
TITLE OF INVENTION: CCK-2, A No. 5677144el Receptor Tyrosine Kinase  
NUMBER OF SEQUENCES: 43  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pennie & Edmonds  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036-2711  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/336,343A  
FILING DATE: 08-NOV-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Coruzzi, Laura A.  
REGISTRATION NUMBER: 30,742  
REFERENCE/DOCKET NUMBER: 7683-065  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-9741/8864  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 18:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 50 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
US-08-336-343A-18

Query Match 73.9%; Score 34; DB 1; Length 50;  
Best Local Similarity 100.0%; Pred. No. 16;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GVCYVD 7  
1|||||

Db 45 GVCVOD 50

RESULT 10  
US-08-701-191A-25  
Sequence 25, Application US/08701191A  
Patent No. 5942428  
GENERAL INFORMATION:  
APPLICANT: Moosa Mohammadi, Joseph Schlessinger,  
APPLICANT: and Steven R. Hubbard  
TITLE OF INVENTION: CRYSTALS OF THE TYROSINE KINASE DOMAIN  
TITLE OF INVENTION: OF NON-INSULIN RECEPTOR TYROSINE KINASE  
NUMBER OF SEQUENCES: 41  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Lyon & Lyon  
STREET: 633 West Fifth Street  
CITY: Los Angeles  
STATE: California  
COUNTRY: U.S.A.  
ZIP: 90071-2066  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
MEDIUM TYPE: storage  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: IBM P.C. DOS 5.0  
SOFTWARE: FASTSEQ for Windows 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/701.191A  
FILING DATE: August 21, 1996  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Warburg, Richard J.  
REGISTRATION NUMBER: 32,327  
REFERENCE/DOCKET NUMBER: 227/088  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (213) 489-1600  
TELEFAX: (213) 955-0440  
TELEX: 67-3510  
INFORMATION FOR SEQ ID NO: 25:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 317 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-701-191A-25

Query Match 73.9%; Score 34; DB 2; Length 317;  
Best Local Similarity 100.0%; Pred. No. 1e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GVCVOD 7  
Db 93 GVCVOD 98

RESULT 11  
US-08-836-567-6  
Sequence 6, Application US/08836567  
Patent No. 6130367  
GENERAL INFORMATION:  
APPLICANT: Rossmann, Jens  
APPLICANT: Sprlinger, Franziska  
APPLICANT: Abel, Gernot  
TITLE OF INVENTION: DNA MOLECULES THAT CODE FOR ENZYMES  
TITLE OF INVENTION: INVOLVED IN STARCH SYNTHESIS VECTORS BACTERIA TRANSGENIC  
TITLE OF INVENTION: PLANT CELLS AND PLANTS CONTAINING SAID MOLECULES  
NUMBER OF SEQUENCES: 17

CORRESPONDENCE ADDRESS:  
ADDRESSEE: FISH & NEAVE  
STREET: 1251 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10020  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/836,567  
FILING DATE: 24-JUL-1997  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/EP95/04415  
FILING DATE: 09-NOV-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: DE P 44 41 408.0  
FILING DATE: 10-NOV-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Haley Jr., James F.  
REGISTRATION NUMBER: 27,794  
REFERENCE/DOCKET NUMBER: Agrevo-4  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-596-9090  
TELEFAX: 212-596-9090  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 558 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-836-567-6

Query Match 73.9%; Score 34; DB 4; Length 558;  
Best Local Similarity 75.0%; Pred. No. 1.8e+02;  
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GVCVODG 8  
Db 190 GVCVODG 197

RESULT 12  
US-08-941-445A-9  
Sequence 9, Application US/08941445A  
Patent No. 6107060  
GENERAL INFORMATION:  
APPLICANT: Keeling, Peter  
APPLICANT: Guan, Hanning  
TITLE OF INVENTION: Starch Encapsulation  
NUMBER OF SEQUENCES: 37  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Greenlee, Winner and Sullivan, P.C.  
STREET: 5370 Manhattan Circle  
CITY: Boulder  
STATE: CO  
COUNTRY: US  
ZIP: 80503  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/941,445A  
FILING DATE: 30-SEP-1997  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 60/026,855  
FILING DATE: 30-SEP-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Winnet, Ellen P  
REGISTRATION NUMBER: 28,547  
REFERENCE/DOCKET NUMBER: 89-97  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (303) 459-8080  
TELEFAX: (303) 499-8089  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 669 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-941-445A-9

Query Match  
Best Local Similarity 73.9%; Score 34; DB 3; Length 669;  
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 GGCVCVDG 8  
||| |  
Db 303 GGCVCYGDG 310

RESULT 13  
US-08-836-567-8  
Sequence 8, Application US/08836567  
Patent No. 6130367  
GENERAL INFORMATION:  
APPLICANT: Kossman, Jens  
APPLICANT: Springer, Franziska  
APPLICANT: Abel, Gernot  
TITLE OF INVENTION: DNA MOLECULES THAT CODE FOR ENZYMES  
INVOLVED IN STARCH SYNTHESIS VECTORS BACTERIA TRANSGENIC  
NUMBER OF INVENTION: PLANT CELLS AND PLANTS CONTAINING SAID MOLECULES  
NUMBER OF SEQUENCES: 17  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: FISH & NEAVE  
STREET: 1251 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10020  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/836,567  
FILING DATE: 24-JUL-1997  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/EP95/04415  
FILING DATE: 09-NOV-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: DE P 44 41 408.0  
FILING DATE: 10-NOV-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Haley Jr., James F.  
REGISTRATION NUMBER: 27,794  
REFERENCE/DOCKET NUMBER: Agrevo-4  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-596-9000  
TELEFAX: 212-596-9090  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 767 amino acids  
TYPE: amino acid  
TOPOLOGY: linear

MOLECULE TYPE: protein  
US-08-836-567-8

Query Match  
Best Local Similarity 73.9%; Score 34; DB 4; Length 767;  
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 GGCVCVDG 8  
||| |  
Db 399 GGCVCYGDG 406

RESULT 14  
US-08-445-640-4  
Sequence 4, Application US/08445640  
Patent No. 5709858  
GENERAL INFORMATION:  
APPLICANT: Godowski, Paul J.  
APPLICANT: Mark, Melanie R.  
APPLICANT: Scadden, David T.  
APPLICANT: Baker, Kevin P.  
APPLICANT: Baron, Will F.  
TITLE OF INVENTION: Protein Tyrosine Kinases  
NUMBER OF SEQUENCES: 35  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genentech, Inc.  
STREET: 460 Point San Bruno Blvd  
CITY: South San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94080

COMPUTER READABLE FORM:  
MEDIUM TYPE: 5.25 inch, 360 kb floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: patin (Genentech)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/445,640  
FILING DATE: 22-MAY-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/170558  
FILING DATE: 20-DEC-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/157563  
FILING DATE: 23-NOV-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Hasak, Janet E.  
REGISTRATION NUMBER: 28,616  
REFERENCE/DOCKET NUMBER: 854C2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415/225-1896  
TELEFAX: 415/952-9881  
TELEX: 910/371-7168  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 913 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
US-08-445-640-4

Query Match  
Best Local Similarity 73.9%; Score 34; DB 1; Length 913;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GGCVD 7  
||| |  
Db 689 GGCYVD 694

RESULT 15

US-08-170-558-4  
; Sequence 4, Application US/08170558  
; Patent No. 6001621  
; GENERAL INFORMATION:  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Mark, Melanie R.  
; APPLICANT: Scadden, David T.  
; APPLICANT: Baker, Kevin P.  
; APPLICANT: Baron, Will F.  
; TITLE OF INVENTION: Protein Tyrosine Kinases  
; NUMBER OF SEQUENCES: 35  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genentech, Inc.  
; STREET: 460 Point San Bruno Blvd  
; CITY: South San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94080  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: palin (Genentech)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/170,558  
; FILING DATE: 20-DEC-1993  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/157563  
; FILING DATE: 23-NOV-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Hasak, Janet E.  
; REGISTRATION NUMBER: 28,616  
; REFERENCE/DOCKET NUMBER: 854C1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415/225-1896  
; TELEFAX: 415/352-9881  
; TELEX: 910/371-7168  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 913 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; US-08-170-558-4

Query Match 73.9%; Score 34; DB 3; Length 913;  
Best Local Similarity 100.0%; Pred. No. 2.9e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GVCVOD 7  
DB 689 GVCVOD 694

Search completed: June 13, 2001, 14:16:35  
Job time: 496 sec

*This Page Blank (uspto)*

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 13, 2001, 14:14:37 ; Search time 229.28 Seconds

(without alignments)  
1.995 Million cell updates/sec

Title: PCT-US01-05825A-15

Perfect score: 42

Sequence: 1 GGVLYQPG 8

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 390729 seqs, 57163235 residues

Total number of hits satisfying chosen parameters: 390729

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A.Geneseq-0401.\*  
1: /SID56/gcgdata/geneseq/AA1980.DAT.\*  
2: /SID56/gcgdata/geneseq/AA1981.DAT.\*  
3: /SID56/gcgdata/geneseq/AA1982.DAT.\*  
4: /SID56/gcgdata/geneseq/AA1983.DAT.\*  
5: /SID56/gcgdata/geneseq/AA1984.DAT.\*  
6: /SID56/gcgdata/geneseq/AA1985.DAT.\*  
7: /SID56/gcgdata/geneseq/AA1986.DAT.\*  
8: /SID56/gcgdata/geneseq/AA1987.DAT.\*  
9: /SID56/gcgdata/geneseq/AA1988.DAT.\*  
10: /SID56/gcgdata/geneseq/AA1989.DAT.\*  
11: /SID56/gcgdata/geneseq/AA1990.DAT.\*  
12: /SID56/gcgdata/geneseq/AA1991.DAT.\*  
13: /SID56/gcgdata/geneseq/AA1992.DAT.\*  
14: /SID56/gcgdata/geneseq/AA1993.DAT.\*  
15: /SID56/gcgdata/geneseq/AA1994.DAT.\*  
16: /SID56/gcgdata/geneseq/AA1995.DAT.\*  
17: /SID56/gcgdata/geneseq/AA1996.DAT.\*  
18: /SID56/gcgdata/geneseq/AA1997.DAT.\*  
19: /SID56/gcgdata/geneseq/AA1998.DAT.\*  
20: /SID56/gcgdata/geneseq/AA1999.DAT.\*  
21: /SID56/gcgdata/geneseq/AA2000.DAT.\*  
22: /SID56/gcgdata/geneseq/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	42	100.0	8	21	Y84661
2	42	100.0	8	21	Y79119
3	42	100.0	20	20	W94489
4	42	100.0	20	21	Y79133
5	42	100.0	25	15	R52775
6	42	100.0	25	15	R52793
7	39	92.9	8	21	Y79123
8	39	92.9	98	21	B40074
9	39	92.9	118	16	R66331
10	39	92.9	124	19	W80816
11	38	90.5	118	16	R66314

12	38	90.5	119	19	W69322
13	38	90.5	119	20	Y32832
14	38	90.5	119	20	W73503
15	38	90.5	120	17	W00240
16	38	90.5	123	15	R54811
17	37	88.1	8	21	Y79117
18	37	88.1	18	19	Y84661
19	37	88.1	20	21	Y84660
20	37	88.1	121	21	B53648
21	36	85.7	118	20	Y15401
22	36	85.7	225	20	Y15409
23	36	85.7	239	20	Y15408
24	36	85.7	247	9	P80155
25	36	85.7	247	9	P80156
26	36	85.7	466	5	P40032
27	36	85.7	500	21	B07654
28	35	83.3	8	21	Y79127
29	35	83.3	13	20	W66096
30	35	83.3	13	20	W66076
31	35	83.3	13	20	W66088
32	35	83.3	18	16	W73953
33	35	83.3	18	16	R82833
34	35	83.3	20	19	W68169
35	35	83.3	20	20	W94487
36	35	83.3	20	21	Y79130
37	35	83.3	27	17	W16581
38	35	83.3	30	17	R87049
39	35	83.3	30	21	Y68811
40	35	83.3	37	17	W00242
41	35	83.3	40	7	P61028
42	35	83.3	62	16	R76969
43	35	83.3	73	21	Y64774
44	35	83.3	84	21	Y64925
45	35	83.3	85	21	Y64735

#### ALIGNMENTS

RESULT 1	
ID: Y84661	standard; Protein: 8 AA.
XX	
AC	Y84661;
XX	
DT	25-JUL-2000 (first entry)
XX	
DE	Peptide antagonist F21/O of zonula occludens toxin (zot) polypeptide.
XX	
KW	Human; zot; zonula occludens toxin; zonulin; antigen presenting cell;
KW	APC; lymphocyte proliferation; antigen; auto-immune disorder;
KW	immune-related disorder; immune system rejection; multiple sclerosis;
KW	organ transplantation; inflammatory disease; allergic disease;
KW	rheumatoid arthritis; insulin dependent diabetes mellitus;
KW	celiac disease; Sjogren's syndrome; systemic lupus erythematosus;
KW	auto-immune thyroiditis; idiopathic thrombocytopenic purpura;
KW	hemolytic anemia; Grave's disease; Addison disease; autoimmune orchitis;
KW	pernicious anemia; vasculitis; autoimmune coagulopathy; polyositis;
KW	myasthenia gravis; polyneuritis; pemphigus; rheumatic carditis;
KW	dermatomyositis; scleroderma; asthma; psoriasis; eczematous dermatitis;
KW	Kaposi's sarcoma; inflammatory bowel disease; proliferative disorder;
KW	PCR primer; ss.
XX	
XX	Synthetic.
OS	
XX	
PN	W0200015252-AA.
XX	
XX	23-MAR-2000.
PD	
XX	
PF	09-SEP-1999; 99WO-US18842.
XX	
PR	14-SEP-1998; 98US-0100266.
XX	

15D3 antibody heav  
15D3 VH chain prot  
Antibody 15D3 heav  
EGF receptor chime  
SPA-reactive VH re  
Peptide antagonist  
Complex scupa/supa  
N-terminal sequenc  
Human colon cancer  
Protein encoded by  
Anti-HCG anti-RR6  
Anti-HCG anti-RR6  
FRS and CDRs from  
Biosynthetic antib  
Gamma anti-carcino  
Llama antibody fra  
Peptide antagonist  
Peptide from human  
Peptide from human  
Peptide from human  
IgG heavy chain fr  
N-terminal of C179  
Human IgG heavy ch  
Human adult heart  
Human adult heart  
Anti-RSV F glycop  
Human group III he  
EGF receptor chime  
H-chain variable r  
HSV-neutralising a  
Human 5' EST relat  
Human 5' EST relat  
Human 5' EST relat

PA (UYMA-) UNIV MARYLAND BALTIMORE.  
 XX Fasano A, Szelein MB, Lu R, Tanner MK;  
 XX WPI: 2000-271257/23.  
 DR  
 XX  
 PT Suppression of antigen presenting cell mediated lymphocyte  
 PT proliferation, by administering a Zot-related immunoregulator useful  
 PT for treating immune-related disorders, immune system rejection  
 PT subsequent to tissue or organ transplantation  
 PS  
 XX Example 7; Page 59; 95pp: English.  
 CC The present sequence represents a peptide antagonist of zonula occludens  
 CC toxin (Zot) polypeptide. The specification describes a method of  
 CC suppressing antigen presenting cell (APC)-mediated lymphocyte  
 CC proliferation in a mammalian host pre-exposed to a particular antigen.  
 CC The method comprises administering to the host an effective amount of a  
 CC Zot-related immunoregulator selected from Zot (zonula occludens toxin) or  
 CC zonulin, the amount effective to down-regulate the activity of the APC.  
 CC The method can be used to down-regulate APC-mediated lymphocyte  
 CC proliferation in mammalian hosts suffering from auto-immune or  
 CC immune-related disorders, immune system rejection subsequent to tissue  
 CC or organ transplantation, or inflammatory or allergic diseases. The  
 CC autoimmune or immune related disorders include multiple sclerosis, the  
 CC rheumatoid arthritis, insulin dependent diabetes mellitus, celiac  
 CC disease, Sjogren's syndrome, systemic lupus erythematosus, auto-immune  
 CC thyroiditis, idiopathic thrombocytopenic purpura, hemolytic anemia,  
 CC Grave's disease, Addison disease, autoimmune orchitis, pernicious  
 CC anemia, vasculitis, autoimmune coagulopathies, myasthenia gravis,  
 CC polynarthritis, pemphigus, rheumatic carditis, polyarthritis,  
 CC dermatomyositis, and scleroderma. The inflammatory or allergic disease  
 CC or disorder is selected from asthma, psoriasis, eczematous dermatitis,  
 CC Karpov's sarcoma, multiple sclerosis, inflammatory bowel disease,  
 CC proliferative disorders of smooth muscle cells, and inflammatory  
 CC conditions associated with mycotic, viral, parasitic, or bacterial  
 CC infections.  
 CC  
 SO Sequence 8 AA:  
 OY 1 GGVLVQPG 8  
 Db 1 ggvlvqpg 8  
 Query Match 100.0%; Score 42; DB 21; Length 8;  
 Best Local Similarity 100.0%; Pred. No. 3.2e+05;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 RESULT 2  
 Y79119  
 ID Y79119 standard; Peptide; 8 AA.  
 XX  
 AC Y79119;  
 DT  
 DT 05-JUN-2000 (first entry)  
 XX  
 DE Peptide antagonist of zonulin.  
 XX  
 KW Zonulin; antagonist; zonula occludens toxin receptor; human;  
 KW blood-brain barrier; antiinflammatory; cerebroprotective;  
 KW neuroprotective; dermatological; antitumor; antiviral;  
 KW antibacterial; cytostatic; anti-HIV; vulnerary; antiallergic;  
 KW hypotensive; immunosuppressive; antiparasitic; vasotropic;  
 KW gastrointestinal inflammation; therapy.  
 XX  
 OS Homo sapiens.  
 OS  
 PN W0200007609-A1.  
 XX  
 PD 17-FEB-2000.  
 XX

PF 28-JUL-1999; 99WO-US16683.  
 XX  
 XX 03-AUG-1998; 98US-0127815.  
 FR  
 XX  
 XX (UYMA-) UNIV MARYLAND BALTIMORE.  
 PA  
 XX Fasano A;  
 PI WPI: 2000-205565/18.  
 DR  
 XX  
 PT New peptide antagonist of zonulin useful as antiinflammatory agent for  
 PT treating cerebral ischemia, stroke, cerebral edema, gastritis,  
 PT shigellosis, viral gastroenteritis, meningitis, encephalomyelitis  
 PT  
 PS Claim 1; Page 45; 69pp: English.  
 CC This present sequence is that of a peptide antagonist of zonulin,  
 CC corresponding to residues 8-15 of human foetal intestinal zonulin.  
 CC It is one of 25 peptide antagonists (see Y79105-29) of the invention  
 CC that bind to a zonula occludens toxin (ZOT) receptor, yet do not  
 CC physiologically modulate the opening of mammalian tight junctions  
 CC (TJ). The peptide antagonists are based on a common motif of ZOT  
 CC and human zonulins, which is believed to be critical for receptor  
 CC binding. They can be prepared by chemical synthesis or by use of  
 CC recombinant DNA techniques. The peptide antagonists are used as an  
 CC antiinflammatory agents in the treatment of gastrointestinal  
 CC inflammation, where they bind to the ZOT receptor in the intestine  
 CC and yet does not physiologically modulate the opening of TJ in the  
 CC intestine. Gastrointestinal inflammation conditions give rise to  
 CC increased intestinal permeability and the peptide is useful for  
 CC treating intestinal conditions that cause protein losing enteropathy  
 CC caused by infection, e.g. Clostridium difficile infection,  
 CC enterocolitis, shigellosis, viral gastroenteritis, parasite  
 CC infestation, bacterial overgrowth, whipple's disease, diseases with  
 CC mucosal erosion or ulcerations, e.g. gastritis, gastric cancer,  
 CC collagenous colitis, inflammatory bowel disease, diseases marked by  
 CC lymphatic obstruction, e.g. congenital intestinal lymphangiectasia,  
 CC sarcoidosis lymphoma, mesenteric tuberculosis, and after surgical  
 CC correction of congenital heart disease with Fontan's operation,  
 CC mucosal diseases without ulceration, e.g. Menetrier's disease,  
 CC coeliac disease, eosinophilic gastroenteritis, and immune diseases,  
 CC e.g. systemic lupus erythematosus or food allergies, primarily to  
 CC milk.  
 CC  
 SO Sequence 8 AA:  
 OY 1 GGVLVQPG 8  
 Db 1 ggvlvqpg 8  
 Query Match 100.0%; Score 42; DB 21; Length 8;  
 Best Local Similarity 100.0%; Pred. No. 3.2e+05;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 RESULT 3  
 W94489  
 ID W94489 standard; peptide; 20 AA.  
 XX  
 AC W94489;  
 DT  
 DT 21-APR-1999 (first entry)  
 XX  
 DE Human foetal intestine zonulin N-terminal peptide.  
 XX  
 KW Zonulin; mammalian tight junction; zonula occludens toxin; ZOT;  
 KW Vibrio cholerae; vaccine; cholera toxin; polyclonal antibody;  
 KW intestinal mucosa; nasal mucosa; blood brain barrier.  
 XX  
 OS Homo sapiens.  
 OS  
 PN  
 XX  
 FH Key Location/Qualifiers



```

FT Misc-difference 16 /note= "unspecified"
XX
XX
XX WO9852415-A1.
XX
XX 26-NOV-1998.
XX
XX 28-APR-1998; 98WO-US07636.
XX
XX 21-MAY-1997; 97US-0859931.
XX
XX (UYMA-) UNIV MARYLAND BALTIMORE.
XX
XX Fasano A:
XX
XX WPI: 1999-070123/06.
XX
XX New purified zonulin - which is capable of reversibly opening
XX mammalian tight junctions, used for enhancing the delivery of agents
XX across intestinal and nasal mucosa and blood brain barrier
XX
XX
XX
XX
XX Claim 2; Page 45; 64pp; English.
XX
XX
XX The present invention describes pure zonulin which has an apparent
XX molecular weight of 47 kD, as determined by SDS-PAGE, which is
XX recognized by both anti- $\tau$ au polyclonal antibody and by anti-zonula
XX occludens toxin (ZOT) polyclonal antibody, and is capable of reversibly
XX opening mammalian tight junctions. Zonulin proteins function as
XX physiological modulators of mammalian tight junctions. They can be used
XX for enhancing the absorption of therapeutic agents across tight
XX junctions of intestinal and nasal mucosa and across tight junctions of
XX the blood brain barrier. Zonulin can be used with agents such as drugs,
XX e.g. lidocaine, adenosine, dobutamine, dopamine, epinephrine,
XX norepinephrine, phenoltamine, doxapram, alfentanil, dezocin, nalbuphine,
XX buprenorphine, naloxone, ketorolac, midazolam, propofol, metacurine,
XX mivacurium, succinylcholine, cytarabine, mitomycin doxorubicin,
XX vincristine, vinblastine, methicillin, mezlocillin, piperacillin,
XX cefoxitin, ceftriaxol, cefmetazole and aztreonam, a hormone e.g.
XX testosterone, nandrolone, menotrolone, insulin, urofollitropin,
XX interferon- $\alpha$ , interferon- $\beta$ , interferon- $\gamma$ , interleukin-1
XX (IL-1), IL-2, IL-4, IL-6, polivalent IgG, specific IgG, IgA, or IgM.
XX The proteins can also be used for the production of antibodies which can
XX be used to assay for zonulin in body tissue or fluids, or in affinity-
XX purification of zonulin. The present sequence represents an N-terminal
XX peptide of zonulin.
XX
XX
XX Sequence 20 AA:
XX
XX
XX Query Match 100.0%; Score 42; DB 20; Length 20;
XX Best Local Similarity 100.0%; Pred. No. 0.28;
XX Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0.
XX
XX 1 GCVLVQPG 8
XX |||||
XX |||||
XX 8 ggvlvvpg 15
XX
XX
XX
XX RESULT 4
XX Y79133
XX Y79133 standard; Peptide: 20 AA.
XX
XX Y79133:
XX
XX 05-JUN-2000 (first entry)
XX
XX Human foetal intestine zonulin N-terminal sequence.
XX
XX Zonulin: antagonist; zonula occludens toxin receptor;
XX human; blood-brain barrier; anti-inflammatory;
XX gastrointestinal inflammation; therapy.
XX
XX Homo sapiens
XX

```

XX	Key	Location/Qualifiers
FH	Misc-difference	16
FT	/note= "unidentified residue"	
XX	MO200007609-A1.	
XX	17-FEB-2000.	
XX	28-JUL-1999;	99WO-US16683.
XX	03-AUG-1998;	98US-0127815.
XX	(UTMA-) UNIV MARYLAND BALTIMORE.	
XX	Fasano A;	
XX	WPI: 2000-205565/18.	
XX	New peptide antagonist of zonulin useful as antiinflammatory agent for treating cerebral ischemia, stroke, cerebral edema, gastritis, shigellosis, viral gastroenteritis, meningitis, encephalomyelitis	
XX	Example 3; Fig 6; 69pp; English.	
XX	The present sequence is that of the N-terminal region of foetal human intestinal zonulin. The N-terminal sequences of human adult and foetal zonulins (see Y79130-36) were compared with Vibrio cholerae zonula occludens toxin (ZOT) to identify a common motif thought to be involved in receptor binding. Peptide antagonists (see Y79105-29) based on this motif are useful as antiinflammatory agents for treatment of gastrointestinal inflammation, and for treatment of conditions associated with breakdown of the blood-brain barrier.	
XX	Sequence	20 AA;
XX	SO	
XX	Query Match	100.0%; Score 42; DB 21; Length 20;
XX	Best Local Similarity	100.0%; Pred. No. 0.28;
XX	Matches	8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY	1 GGVLVQPG 8	
DB	8 99VLVQPG 15	
RESULT	5	
RS2775	R52775 standard; Protein; 25 AA.	
XX	AC	
XX	RS2775;	
DT	24-JAN-1995	(first entry)
DE	Murine BRE-3 immunoglobulin heavy chain variable domain N-terminus.	
XX	Immunoglobulin variable domain; primer; polymerase chain reaction; chimeric antibody; human milk fat globule; BRE-3 VL-chain.	
KW	Mus musculus.	
OS	MO9411508-A.	
XX	PN	
XX	26-MAY-1994.	
XX	PD	
XX	15-NOV-1993;	93WO-US11316.
XX	PF	
PR	13-NOV-1992;	92US-0977706.
PR	13-NOV-1992;	92US-0977707.
PR	28-SEP-1993;	93US-0128015.
XX	(CANC-) CANCER RES FUND CONTRA COSTA.	

XX WPI; 1994-183509/22.  
 XX Chimeric human-murine polypeptide(s) specific for human mammary  
 PT fat globule antigen - for imaging, diagnosing and treating  
 PT neoplasia, with less undesirable immunogenic response  
 XX  
 XX Example 11: Page 32; 54pp; English.  
 XX  
 CC Primers J02, J03, J04, J014 and VH1BACK (062740-062744) were all  
 CC used to prepare cDNAs that encode the Bre-3 mouse Ig variable  
 CC domains. The amplified V-regions lacked constant regions so as to  
 CC produce less immunogenic polypeptides. A hybrid polypeptide was  
 CC prepared using human constant regions with the murine V regions.  
 CC The chimeric polypeptide retained the binding affinity of Bre-3  
 CC for human milk fat globule. The amino acid sequences deduced from  
 CC the amplified VL and VH regions were compared to N-terminal  
 CC sequences which had been determined directly (i.e. R52774 and  
 CC R52775). The general agreement between the predicted and the  
 CC determined amino acid sequences that the cloned cDNAs encode Bre-3.  
 CC  
 SQ Sequence 25 AA:

Query Match 100.0%; Score 42; DB 15; Length 25;  
 Best Local Similarity 100.0%; Pred. No. 0.35;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GGVLVQPG 8  
 |||||  
 Db 8 ggvlvqpg 15

RESULT 6  
 ID R52793 standard; Protein: 25 AA.  
 XX R52793:  
 AC 24-JAN-1995 (first entry)  
 DT  
 XX  
 DE Murine Bre-3 immunoglobulin heavy chain variable domain N-terminus.  
 XX  
 KM Immunoglobulin variable domain; primer; polymerase chain reaction;  
 KW chimeric antibody; human milk fat globule; Bre-3 VL-chain.  
 XX  
 OS Mus musculus.  
 XX  
 PN MO9411509-A.  
 PD 26-MAY-1994.  
 PF 16-NOV-1993; 93WO-US11445.  
 PR 16-NOV-1992; 92US-0977696.  
 PR 30-SEP-1993; 93US-0129930.  
 PR 08-OCT-1993; 93US-0134346.  
 XX  
 XX (CANC-) CANCER RES FUND CONTRA COSTA.  
 XX  
 DR WPI; 1994-183510/22.  
 XX  
 PT New analogue peptide(s) comprising antibody variable regions -  
 PT used to develop prods. for use in the detection, diagnosis,  
 PT therapy and prevention of neoplasms  
 XX  
 XX Example 12: Page 53; 54pp; English.  
 XX  
 CC Primers J02, J03, J04, J014 and VH1BACK (062765-062769) were all  
 CC used to prepare cDNAs that encode the Bre-3 mouse Ig variable  
 CC domains. The amplified V-regions lacked constant regions so as to  
 CC produce less immunogenic polypeptides. A hybrid polypeptide was  
 CC prepared using human constant regions with the murine V regions.

CC The chimeric polypeptide retained the binding affinity of Bre-3  
 CC for human milk fat globule. The amino acid sequences deduced from  
 CC the amplified VL and VH regions were compared to N-terminal  
 CC sequences which had been determined directly (i.e. R52792 and  
 CC R52793). The general agreement between the predicted and the  
 CC determined amino acid sequences that the cloned cDNAs encode Bre-3.  
 CC  
 SQ Sequence 25 AA:

Query Match 100.0%; Score 42; DB 15; Length 25;  
 Best Local Similarity 100.0%; Pred. No. 0.35;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GGVLVQPG 8  
 |||||  
 Db 8 ggvlvqpg 15

RESULT 7  
 ID Y79123 standard; Peptide: 8 AA.  
 XX Y79123:  
 AC 05-JUN-2000 (first entry)  
 DT  
 XX  
 DE Peptide antagonist of zonulin.

XX Zonulin; antagonist; zonula occludens toxin receptor;  
 KW blood-brain barrier; antiinflammatory; cerebroprotective;  
 KW neuroprotective; dermatological; antitumor; antiviral;  
 KW antibacterial; cytostatic; anti-HIV; antiviral; antifungal;  
 KW hypotensive; immunosuppressive; antiparasitic; vasotropic;  
 KW gastrointestinal inflammation; therapy.  
 XX  
 OS Synthetic.  
 XX  
 PN WO200007609-A1.  
 PD 17-FEB-2000.  
 PF 28-JUL-1999; 99WO-US16683.  
 PR 03-AUG-1998; 98US-0127815.  
 XX  
 XX (UYMA-) UNIV MARYLAND BALTIMORE.  
 XX  
 PA Fasano A;  
 PI  
 PT  
 XX  
 DR WPI; 2000-205565/18.  
 XX  
 PT New peptide antagonist of zonulin useful as antiinflammatory agent for  
 PT treating cerebral ischemia, stroke, cerebral edema, gastritis,  
 PT shigellosis, viral gastroenteritis, meningitis, encephalomyelitis -  
 XX  
 PS Claim 1: Page 47; 69pp; English.

CC This present sequence is that of a peptide antagonist of zonulin  
 CC (Z). One of 25 such peptides (see Y79105-29) of the invention,  
 CC which bind to a zonula occludens toxin (ZOT) receptor, yet do not  
 CC physiologically modulate the opening of mammalian tight junctions  
 CC (TJ). The peptide antagonists are based on a common motif of ZOT  
 CC and human zonulins, which is believed to be critical for receptor  
 CC binding. They can be prepared by chemical synthesis or by use of  
 CC recombinant DNA techniques. The peptide antagonists are used as an  
 CC antiinflammatory agents in the treatment of gastrointestinal  
 CC inflammation, where they bind to the ZOT receptor in the intestine  
 CC and yet does not physiologically modulate the opening of TJ in the  
 CC intestine. Gastrointestinal inflammation conditions give rise to  
 CC increased intestinal permeability and the peptide is useful for  
 CC treating intestinal conditions that cause protein losing enteropathy  
 CC caused by infection, e.g. Clostridium difficile infection,

CC enterocolitis, shigellosis, viral gastroenteritis, parasite  
 CC infection, bacterial overgrowth, whipple's disease, diseases with  
 CC mucosal erosion or ulcerations, e.g., gastritis, gastric cancer,  
 CC collagenous colitis, inflammatory bowel disease, diseases marked by  
 CC lymphatic obstruction, e.g. congenital intestinal lymphangiectasia,  
 CC sarcoidosis lymphoma, mesenteric tuberculosis, and after surgical  
 CC correction of congenital heart disease with Fontan's operation,  
 CC mucosal diseases without ulceration, e.g. Menetrier's disease,  
 CC coeliac disease, eosinophilic gastroenteritis, and immune diseases,  
 CC e.g. systemic lupus erythematosus or food allergies, primarily to  
 CC milk.  
 XX  
 SO Sequence 8 AA:  
 Query Match 92.9%; Score 39; DB 21; Length 8;  
 Best Local Similarity 87.5%; Pred. No. 3.2e+05;  
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 GGVLVQPG 8  
 ||:|||||  
 Db 1 gglivqpg 8  
 RESULT 8  
 B40074  
 ID B40074 standard; Protein; 98 AA.  
 AC B40074;  
 XX  
 DT 05-FEB-2001 (first entry)  
 DE Anti-IL12 antibody H chain V region amino acid sequence SEQ ID 600.  
 XX  
 KW Human; neutralising antibody; interleukin-12; IL-12; antiinflammatory;  
 KW complementarily determining region; CDR; antihemmatic; antiarthritic;  
 KW antileptotic; neuroprotective; antipsoriatic; antiaesthetic; cardiant;  
 KW antiparasitic; antibacterial; immunosuppressive; Crohn's disease;  
 KW multiple sclerosis; rheumatoid arthritis.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200056772-A1.  
 XX  
 PD 28-SEP-2000.  
 XX  
 PF 24-MAR-2000; 2000WO-US07946.  
 XX  
 PR 25-MAR-1999; 99US-0126603.  
 XX  
 PA (BADI ) BASF AG.  
 PA (GEMV ) GENETICS INST INC.  
 XX  
 PI Salfeld JS, Roguska M, Paskind M, Banerjee S, Tracey DE, White M;  
 PI Kaymakcalan Z, Labkovsky B, Sakorafas P, Friedrich S, Myles A;  
 PI Veldman GM, Vennurini A, Warne NW, Widom A, Elyin JG, Duncan AR;  
 PI Derbyshire EJ, Carmen S, Smith S, Holtet TL, Du Fou SL;  
 XX  
 DR WPI: 2000-638250/61.  
 XX  
 PT New human antibody specific for human interleukin-12 (IL-12) used to  
 PT treat disorders characterized by aberrant IL-12 expression e.g. Crohn's  
 PT disease and multiple sclerosis -  
 XX  
 PS Claim 75; Page 121; 377pp; English.  
 XX  
 CC This invention relates to a new human antibody specific for human  
 CC interleukin-12 (IL-12). The invention also includes antigen binding  
 CC portions that bind to IL-12. Sequences B39485-B39516 represent human  
 CC anti-IL-12 antibody heavy and light chain complementarily determining  
 CC region (CDR) amino acid sequences, and also includes variable region  
 CC amino acid sequences. Other variable region amino acid sequences are  
 CC given in B39517-B39560 and B40068-B40149. Sequences B39561-B39771

CC represent anti-IL-12 CDR3 related amino acid sequences, B39772-B40063  
 CC represent other CDR sequences. Light chain CDR3 consensus sequences are  
 CC given in B40064-B40067. Primers used in the identification and  
 CC construction of the antibodies of the invention are given in  
 CC C61062-C61071. The antibody of the invention is a neutralising antibody  
 CC and has antihemmatic; antiarthritic; antileptotic; antiinflammatory;  
 CC neuroprotective; antipsoriatic; antiaesthetic; cardiant; antiparasitic;  
 CC antibacterial and immunosuppressive activity. The antibodies or  
 CC antigen-binding fragments are useful in the treatment of disorders  
 CC associated with detrimental release of human IL-12, especially Crohn's  
 CC disease, multiple sclerosis and rheumatoid arthritis. They can also be  
 CC used in the manufacture of a pharmaceutical composition to treat human  
 CC IL-12 disorders.  
 XX  
 SO Sequence 98 AA:  
 Query Match 92.9%; Score 39; DB 21; Length 98;  
 Best Local Similarity 87.5%; Pred. No. 4.9;  
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 GGVLVQPG 8  
 ||:|||||  
 Db 8 ggvvvpqg 15  
 RESULT 9  
 R66331  
 ID R66331 standard; Protein; 118 AA.  
 AC R66331;  
 XX  
 DT 04-AUG-1995 (first entry)  
 DE Human immunoglobulin variable heavy chain #37.  
 XX  
 KW Primer: PCR: amplify; human; immunoglobulin; variable; heavy chain;  
 KW cosmid; placenta; vector; pUB81; E.coli; mammalian.  
 KW  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO9426895-A.  
 XX  
 PD 24-NOV-1994.  
 XX  
 PF 10-MAY-1993; 93WO-JP00603.  
 XX  
 PR 10-MAY-1993; 93WO-JP00603.  
 XX  
 PA (NISB ) JAPAN TOBACCO INC.  
 XX  
 PI Honjo T, Matsuda F;  
 XX  
 DR WPI: 1995-006791/01.  
 DR N-PSDB: Q78981.  
 XX  
 PT DNA fragment comprising human immunoglobulin Vh genes - for the  
 PT production of human immunoglobulin in mammalian hosts  
 XX  
 PS Claim 50; Page 85-87; 130pp; Japanese.  
 XX  
 CC Protein sequences (R66295-51) are novel human immunoglobulin heavy chain  
 CC sequences encoded by novel isolated genes. The genes (Q78939-79002) were  
 CC isolated and cloned from a series of cosmid constructs: Y202; Y103; Y21;  
 CC Y6; Y24; 3-31; M84; M18 and M131, by PCR amplification using primers  
 CC Q78917-38. The genes are subdivided into 5 families of Vh genes. The  
 CC fragments cover a region of 800 kb. The DNA fragments were isolated from  
 CC high molecular weight DNA from human placenta. The DNA was partially  
 CC digested with XbaI restriction enzyme. The fragments were separated by  
 CC gel electrophoresis and 35-45 kb fractions were collected. The fragments  
 CC were ligated with ClaI-digested cosmid vector pUB81. The ligation  
 CC products were in vitro packed and infected into E.coli 490A. The  
 CC fragments were then subcloned by colony hybridisation. The Vh genes and

CC the DNA fragments encoding them are useful in producing human  
 CC immunoglobulin in mammalian hosts.  
 XX  
 SQ Sequence 118 AA;

Query Match 92.9%; Score 39; DB 16; Length 118;  
 Best Local Similarity 87.5%; Pred. No. 5.9;  
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GGVLYOPG 8  
 |||:||||  
 Db 27 ggvvvpqg 34

RESULT 10  
 W80816  
 ID W80816 standard; Protein; 124 AA.  
 XX  
 AC W80816;  
 XX  
 DT 16-FEB-1999 (first entry)  
 XX

DE Amino acid sequence of human D7.2 heavy chain variable region.

KW Human; D7.2 heavy chain variable region; receptor; antigen;  
 KM tumour; auto-immune disease; graft rejection; allergy;  
 XX inflammatory disease; endocrine disease; degenerative disease.

OS Homo sapiens.

PN W09846645-A2.

PD 22-OCT-1998.

PF 14-APR-1998; 98WO-EP02180.

PR 14-APR-1997; 97EP-0106109.

PA (KUFE/) KUFER P.

XX (RAUM/) RAUM T.

PI Kufer P, Raum T;

XX

DR WPI; 1998-594564/50.

DR N-PSDB; V68538.

XX

PT Production of anti-human antigen receptors - by selecting a

PT combination of functionally rearranged VH and VL immunoglobulin

PT chains expressed from a recombinant vector

XX

PS Claim 9; Fig 8; 84pp; English.

XX

CC This is the amino acid sequence of the human D7.2 heavy chain

CC variable region, used in the method of the invention, for providing

CC receptors that can be used for targeting antigens in humans without

CC being immunogenic themselves. Such receptors can be used for treating

CC diseases such as tumours or auto-immune diseases, graft rejection

CC after transplantation; infectious diseases by targeting cellular

CC receptors as well as allergic, inflammatory, endocrine and

CC degenerative diseases by targeting key molecules involved in the

CC pathological process.

XX

SQ Sequence 124 AA;

OY

Db

1 GGVLYOPG 8

|||:||||

8 ggvvvpqg 15

92.9%; Score 39; DB 19; Length 124;

Best Local Similarity 87.5%; Pred. No. 6.2;

Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

RESULT 11

R66314

XX R66314 standard; Protein; 118 AA.

AC R66314;

XX

DT 03-AUG-1995 (first entry)

XX

DE Human immunoglobulin variable heavy chain #20.

XX

KW Primer; PCR; amplify; human; immunoglobulin; variable; heavy chain;

KW cosmid; placenta; vector; pJB81; E.coli; mammalian.

XX

OS Homo sapiens.

PN W09426895-A.

PD 24-NOV-1994.

PF 10-MAY-1993; 93WO-JP00603.

PR 10-MAY-1993; 93WO-JP00603.

XX

PA (NISH ) JAPAN TOBACCO INC.

XX

PI Honjo T, Matsuda F;

XX

DR WPI; 1995-006791/01.

DR N-PSDB; Q78960.

XX

PT DNA fragment comprising human immunoglobulin Vh genes - for the

PT production of human immunoglobulin in mammalian hosts

XX

PS Claim 31; Page 59-60; 130pp; Japanese.

XX

CC Protein sequences (R66295-51) are novel human immunoglobulin heavy chain

CC sequences encoded by novel isolated genes. The genes (Q78939-79002) were

CC isolated and cloned from a series of cosmid constructs: Y202; Y103; Y21;

CC Y6; Y24; 3-31; M84; M18 and M131, by PCR amplification using primers

CC Q78917-38. The genes are subdivided into 5 families of Vh genes. The

CC fragments cover a region of 800 kd. The DNA fragments were isolated from

CC high molecular weight DNA from human placenta. The DNA was partially

CC digested with TaqI restriction enzyme. The fragments were separated by

CC gel electrophoresis and 35-45 kb fractions were collected. The fragments

CC were ligated with ClaI-digested cosmid vector pJB81. The ligation

CC products were in vitro packed and infected into E.coli 430A. The

CC fragments were then subcloned by colony hybridisation. The Vh genes and

CC the DNA fragments encoding them are useful in producing human

CC immunoglobulin in mammalian hosts.

XX

SQ Sequence 118 AA;

OY

Db

1 GGVLYOPG 8

|||:||||

27 ggalvvpqg 34

90.5%; Score 38; DB 16; Length 118;

Best Local Similarity 87.5%; Pred. No. 8.9;

Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

RESULT 12

W69322

ID W69322 standard; Protein; 119 AA.

XX

AC W69322;

XX

DT 19-NOV-1998 (first entry)

XX

DE 15D3 antibody heavy chain.



W73503  
ID W73503 standard; Protein; 119 AA.  
XX  
AC W73503;  
XX  
DT 26-FEB-1999 (first entry)  
XX  
DE Antibody 15D3 heavy chain protein sequence.  
XX  
KW Antibody; 15D3; human; tumour cell; multiple drug resistance; cancer;  
XX immunotherapy; heavy chain.  
XX  
OS Homo sapiens.  
XX  
PN US5849877-A.  
XX  
PD 15-DEC-1998.  
XX  
PF 07-JUN-1995; 95US-0483199.  
XX  
PR 07-JUN-1995; 95US-0483199.  
XX 29-OCT-1990; 90US-0605399.  
PR 22-OCT-1993; 93US-0141375.  
PR 17-OCT-1994; 94US-0323566.  
XX  
PA (CHIR ) CHIRON CORP.  
XX  
PI Ring DB;  
XX  
DR WPI: 1999-069806/06.  
DR N-PSDB; V08933.  
XX  
PT Monomeric and dimeric poly:peptide(s) - which bind to human tumour  
XX cell displaying multiple drug resistance  
XX  
PS Claim 1: Fig 1; 26pp; English.  
XX  
CC This sequence represents the heavy chain of the antibody produced by  
XX the hybridoma 15D3. The protein is an example of a monomeric  
CC polypeptide of the invention that binds to a human tumour cell displaying  
CC multiple drug resistance. The polypeptides are used in specific binding  
CC assays, affinity purification, drug or toxin targeting, imaging or  
CC immunotherapy of various cancers.  
XX  
SQ Sequence 119 AA;  
XX

Query Match 90.5%; Score 38; DB 20; Length 119;  
Best Local Similarity 87.5%; Pred. No. 9;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GGVLVQPG 8  
DB 8 ggvlvtrpg 15

RESULT 15  
W00240  
ID W00240 standard; Protein; 120 AA.  
XX  
AC W00240;  
XX  
DT 22-NOV-1996 (first entry)  
XX  
DE EGF receptor chimeric Mab chMint5 VH chain.  
XX  
KW Mouse-human chimeric antibody; monoclonal antibody; chMint5;  
XX epidermal growth factor receptor; EGF-R; diagnosis; therapy;  
XX immunotoxin; immunocytokine; tumour; cancer.  
XX  
OS Mus musculus.  
XX  
FH Key Location/Qualifiers

FT Region 31..35  
FT /label= CDR1  
FT Region 51..66  
FT /label= CDR2  
FT Region 99..109  
FT /label= CDR3  
XX  
XX W09627010-A1.  
XX  
XX PD 06-SEP-1996.  
XX  
XX PF 01-MAR-1996; 96WO-EP00805.  
XX  
XX PR 01-MAR-1995; 95IT-OFI0036.  
XX  
XX (ITUY-) ITAL MIN UNIV RICERCA SCI & TECNOLOGICA.  
XX  
XX Anastasi AM, Colnaghi MI, De Santis R, Di Massimo AM;  
XX Ferrer Marsal C, Mele A;  
XX WPI: 1996-412776/41.  
XX DR N-PSDB; T33445.  
XX  
XX PT Murine/human chimeric monoclonal antibody, chMint5 specific for  
XX EGF-R - shows a lower immunogenicity when administered to humans  
XX  
XX PS Claim 7: Page 15; 28pp; English.  
XX  
XX CC The amino acid sequence (W00240) of the heavy chain variable region  
XX (VH) of the epidermal growth factor receptor (EGF-R)-specific mouse-  
CC human chimeric antibody chMint5 was deduced from a cDNA clone  
CC (T33445) obtd. by PCR amplification of murine Mint5 hybridoma DSM  
CC ACC2150 cDNA. chMint5 comprises Mint5 VH and VL regions fused  
CC to human C-gamma1 and Ck regions. Constructs were expressed in  
CC CHO cell transfectants. chMint5 shows lower immunogenicity than  
CC Mint5 when administered to humans. It can be used in diagnostic  
CC assays or used to produce immunotoxins or immunocytokines useful  
CC for tumour therapy.  
XX  
SQ Sequence 120 AA;  
XX

Query Match 90.5%; Score 38; DB 17; Length 120;  
Best Local Similarity 87.5%; Pred. No. 9;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GGVLVQPG 8  
DB 8 ggalvqpg 15

Search completed: June 13, 2001, 14:14:38  
Job time: 379 sec



*This Page Blank (uspto)*



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 13, 2001, 14:10:45 ; Search time 130.61 Seconds

(without alignments)  
4.209 Million cell updates/sec

Title: PCT-US01-05825a-15

Perfect score: 42

Sequence: 1 GGVIVQPG 8

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 198801 seqs, 68722935 residues

Total number of hits satisfying chosen parameters: 198801

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	39	92.9	98	2 S26929	Ig heavy chain V r
2	39	92.9	120	2 S44111	Ig heavy chain V-D
3	39	92.9	427	2 D83347	probable aminotran
4	38	90.5	116	1 A1H0ZP	Ig heavy chain V-I
5	36	85.7	96	2 S54854	Ig heavy chain V r
6	36	85.7	110	2 S36282	Ig heavy chain V r
7	36	85.7	251	2 H75301	hypothetical prote
8	36	85.7	263	2 T35032	probable hydroxyla
9	36	85.7	264	2 S12027	methionyl aminopep
10	35	83.3	32	2 F31485	Ig heavy chain V r
11	35	83.3	36	2 D31485	Ig heavy chain V r
12	35	83.3	38	2 S33402	Ig heavy chain V r
13	35	83.3	59	2 A27606	Ig heavy chain V-a
14	35	83.3	59	2 S36361	Ig heavy chain V r
15	35	83.3	60	2 S36382	Ig heavy chain V r
16	35	83.3	82	2 C34964	Ig heavy chain pre
17	35	83.3	94	2 D25913	Ig heavy chain V r
18	35	83.3	94	2 I67528	CD33 antigen homol
19	35	83.3	95	2 I67527	CD33 antigen homol
20	35	83.3	97	2 PH0872	Ig heavy chain V r
21	35	83.3	97	2 S26885	Ig heavy chain V r
22	35	83.3	97	2 S26886	Ig heavy chain V r
23	35	83.3	97	2 S26890	Ig heavy chain V r
24	35	83.3	97	2 S46462	Ig heavy chain V r
25	35	83.3	97	2 S26885	Ig heavy chain V r
26	35	83.3	97	2 S54855	Ig heavy chain V r
27	35	83.3	98	2 PL0121	Ig heavy chain V-I
28	35	83.3	98	2 PH0874	Ig heavy chain V r
29	35	83.3	98	2 PL0123	Ig heavy chain V-I

30	35	83.3	98	2 S26896	Ig heavy chain V r
31	35	83.3	98	2 S29545	Ig heavy chain V r
32	35	83.3	98	2 S26927	Ig heavy chain V r
33	35	83.3	98	2 S26932	Ig heavy chain V r
34	35	83.3	98	2 S26891	Ig heavy chain V r
35	35	83.3	98	2 S26894	Ig heavy chain V r
36	35	83.3	98	2 S26889	Ig heavy chain V r
37	35	83.3	98	2 S26933	Ig heavy chain V r
38	35	83.3	98	2 S26934	Ig heavy chain V r
39	35	83.3	100	2 PL0122	Ig heavy chain V-I
40	35	83.3	100	2 S69896	Ig heavy chain V r
41	35	83.3	100	2 S26925	Ig heavy chain V r
42	35	83.3	100	2 S26926	Ig heavy chain V r
43	35	83.3	101	2 P00003	Ig heavy chain V r
44	35	83.3	101	2 P00002	Ig heavy chain V r
45	35	83.3	105	2 S38488	Ig heavy chain - h

## ALIGNMENTS

```
RESULT 1
S26929
Ig heavy chain V region (DP-33) - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 22-Nov-1993 #sequence_revision 10-Nov-1995 #text_change 23-Jul-1999
C:Accession: S26929
R:Tomlinson, I.M.; Walter, G.; Marks, J.D.; Llewellyn, M.B.; Winter, G.
J. Mol. Biol. 227, 776-798, 1992
A:Title: The repertoire of human germline V(H) sequences reveals about fifty groups o
A:Reference number: S26885; MUID:93021117
A:Accession: S26929
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-98 <TOM>
A:Cross-references: EMBL:Z12335; NID:G32889; PIDN:CAA76205.1; PID:G32890
C:Superfamily: immunoglobulin V region; immunoglobulin homology
C:Keywords: heterotetramer; immunoglobulin
F:15-98/Domain: immunoglobulin homology <IMM>

Query Match 92.9%; Score 39; DB 2; Length 98;
Best Local Similarity 87.5%; Pred. No. 1;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GGVIVQPG 8
Db 8 GGVIVQPG 15

RESULT 2
S44111
Ig heavy chain V-D-J region - human
C:Species: Homo sapiens (man)
C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 23-Jul-1999
C:Accession: S44111
R:Hawkins, R.E.; Zhu, D.; Owecka, M.; Winter, G.; Hamblin, T.J.; Stevenson, F.K.
submitted to the EMBL Data Library, March 1994
A:Description: Idiotypic vaccination against human B-cell lymphoma: rescue of variabl
A:Reference number: S44105
A:Accession: S44111
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-120 <HAN>
A:Cross-references: EMBL:Z31387; NID:G472965; PIDN:CAA83262.1; PID:G940522
C:Superfamily: immunoglobulin V region; immunoglobulin homology
C:Keywords: heterotetramer; immunoglobulin
F:15-98/Domain: immunoglobulin homology <IMM>

Query Match 92.9%; Score 39; DB 2; Length 120;
Best Local Similarity 87.5%; Pred. No. 1.3;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

OY 1 GGVLPDPG 8  
1111111  
DB 8 GGVLPDPG 15

## RESULT 3

Probable aminotransferase Pf2394 [Imported] - Pseudomonas aeruginosa (strain PA01)  
C:Species: Pseudomonas aeruginosa  
C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
C:Accession: D83347  
R:Stover, C.K.; Pham, X.Q.; Ertwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Berman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lam, J.; Lory, S.; Olson, M.V.  
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathogen  
A:Reference number: A82950; MUID:20437337  
A:Accession: D83347  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-427 <STO>  
A:Cross-references: GB:AE004665; GB:AE004091; NID:g9948426; PIDN:AA605782.1; GSPDB:GN001  
C:Genetics:  
A:Gene: PA2394

## Query Match

Best Local Similarity 92.9%; Score 39; DB 2; Length 427;  
Best Local Similarity 87.5%; Pred. No. 4.6;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GGVLPDPG 8  
1111111  
DB 131 GGVLPDPG 138

## RESULT 4

IGHV2P  
Ig heavy chain V-II region (2ap) - human (tentative sequence)  
C:Species: Homo sapiens (man)  
C:Date: 24-Apr-1984 #sequence\_revision 24-Apr-1984 #text\_change 31-Mar-2000  
C:Accession: A02061  
R:Capra, J.D.; Kehoe, J.M.  
Proc. Natl. Acad. Sci. U.S.A. 71, 845-848, 1974  
A:Title: Variable region sequences of five human immunoglobulin heavy chains of the V-II region  
A:Reference number: A93794; MUID:74142702  
A:Accession: A02061  
A:Molecule type: protein  
A:Residues: 1-116 <CAP>  
C:Comment: This chain was isolated from an IgA1 myeloma protein.  
C:Genetics:  
A:Gene: GDB:IGHV@  
A:Cross-references: GDB:128528; OMIM:147070  
A:Map position: 14q32.33-14q32.33  
C:Superfamily: immunoglobulin V region; immunoglobulin homology  
C:Keywords: heterotetramer; immunoglobulin  
F:15-98/Domain: immunoglobulin homology <IMM>  
F:22-96/Disulfide bonds: #status predicted

Query Match 90.5%; Score 38; DB 1; Length 116;  
Best Local Similarity 87.5%; Pred. No. 2;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GGVLPDPG 8  
1111111  
DB 8 GGVLPDPG 15

## RESULT 5

SS4854  
Ig heavy chain V region - human (fragment)

C:Species: Homo sapiens (man)

C:Date: 12-Feb-1998 #sequence\_revision 20-Feb-1998 #text\_change 21-Jan-2000

C:Accession: S54854

R:Tominson, M.; Walter, G.; Cook, X.Y.Z.; Winter, G.

submitted to the EMBL Data Library, November 1992

A:Reference number: S54854

A:Accession: S54854

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-96 <TOM>

A:Cross-references: EMBL:218898; NID:g840776; PIDN:CAA79335.1; PID:g860971

C:Superfamily: immunoglobulin V region; immunoglobulin homology

F:15-96/Domain: immunoglobulin homology <IMM>

## Query Match

Best Local Similarity 85.7%; Score 36; DB 2; Length 96;  
Best Local Similarity 100.0%; Pred. No. 4;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

## RESULT 6

S36282  
Ig heavy chain V region (clone alpha-FOG1-H6) - human (fragment)  
C:Species: Homo sapiens (man)  
C:Date: 03-Feb-1994 #sequence\_revision 03-Feb-1994 #text\_change 23-Jul-1999  
C:Accession: S36282  
R:Griffiths, A.D.; Malmqvist, M.; Marks, J.D.; Bye, J.M.; Embleton, M.J.; McCafferty, E.M.  
EMBO J. 12, 725-734, 1993  
A:Title: Human anti-self antibodies with high specificity from phage display libraries  
A:Reference number: S36256; MUID:93178448  
A:Accession: S36282  
A:Status: preliminary; nucleic acid sequence not shown  
A:Molecule type: mRNA  
A:Residues: 1-110 <GRI>  
A:Cross-references: EMBL:218824; NID:g33111; PIDN:CAA79276.1; PID:g939891  
C:Superfamily: immunoglobulin V region; immunoglobulin homology  
C:Keywords: heterotetramer; immunoglobulin  
F:15-98/Domain: immunoglobulin homology <IMM>

Query Match 85.7%; Score 36; DB 2; Length 110;  
Best Local Similarity 87.5%; Pred. No. 4.6;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GGVLPDPG 8  
1111111  
DB 8 GGVLPDPG 15

## RESULT 7

H75301  
hypothetical protein - Deinococcus radiodurans (strain R1)  
C:Species: Deinococcus radiodurans  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 31-Mar-2000  
C:Accession: H75301  
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.; Shen, M.; Vamathevan, J.U.; Lam, P.; McDonald, L.; Uitterback, T.; Zalewski, C.; Smith, H.O.; Venter, J.C.; Fraser, C.M.  
Science 286, 1571-1577, 1999  
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.  
A:Reference number: A75250; MUID:20036896  
A:Accession: H75301  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-251 <WHI>  
A:Cross-references: GB:AE002054; GB:AE00513; NID:g6460010; PIDN:AA11766.1; PID:g646  
A:Experimental source: strain R1  
C:Genetics:  
A:Gene: DR2216

A:Map position: 1

Query Match 85.7%; Score 36; DB 2; Length 251;  
 Best Local Similarity 75.0%; Pred. No. 11;  
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGVLPQG 8  
 |||||:  
 DB 51 GGVLLLEPG 58

RESULT 8  
 T35032  
 Probable hydroxylase - Streptomyces coelicolor

C:Species: Streptomyces coelicolor  
 C:Date: 05-Nov-1999 #sequence\_revision 05-Nov-1999 #text\_change 21-Jan-2000  
 C:Accession: T35032  
 R:Seeger, S.; Harris, D.; James, K.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.  
 submitted to the EMBL Data Library, June 1999

A:Reference number: 221565  
 A:Accession: T35032  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-263 <SEB>  
 A:Cross-references: EMBL:AL079355; PIDN:CAB45588.1; GSPDB:GN00070; SCOEDB:SC4C6.24C  
 A:Experimental source: strain A3(2)  
 C:Genetics:  
 A:Gene: SCOEDB:SC4C6.24C  
 C:Superfamily: Mycobacterium hypothetical protein RV0911

Query Match 85.7%; Score 36; DB 2; Length 263;  
 Best Local Similarity 100.0%; Pred. No. 11;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGVLPQ 7  
 |||||:  
 DB 92 GGVLPQ 98

RESULT 9  
 S12027  
 methionyl aminopeptidase (EC 3.4.11.18) - Salmonella typhimurium

N:Alternate names: peptidase M  
 C:Species: Salmonella typhimurium  
 C:Date: 21-Nov-1993 #sequence\_revision 10-Nov-1995 #text\_change 22-Jun-1999  
 C:Accession: S12027; S03562  
 R:MOVVA, N.R.; Semon, D.; Meyer, C.; Kawashima, E.; Wingfield, P.; Miller, J.L.; Miller, M.L. Gen. Genet. 223, 345-348, 1990  
 A:Title: Cloning and nucleotide sequence of the Salmonella typhimurium pepm gene.  
 A:Reference number: S12027; MUID:9106851

A:Accession: S12027  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-264 <MOV>  
 A:Cross-references: GB:X55778; MID:947818; PIDN:CA93298.1; PID:947819  
 R:Wingfield, P.; Grader, P.; Tuccatelli, G.; Movva, N.R.; Pelletier, M.; Craig, S.; Rose, Eur. J. Biochem. 180, 23-32, 1989  
 A:Title: Purification and characterization of a methionine-specific aminopeptidase from  
 A:Reference number: S03562; MUID:89210857  
 A:Accession: S03562  
 A:Status: nucleic acid sequence not shown

A:Molecule type: DNA  
 A:Residues: 2-264 <WIN>  
 A>Note: part of this sequence, including the amino end of the mature protein, was confir  
 C:Genetics:  
 A:Gene: pepm  
 C:Superfamily: Escherichia coli methionyl aminopeptidase  
 C:Keywords: alpha-aminoacylpeptide hydrolase  
 F:2-264/Product: methionine aminopeptidase #status experimental <MAT>

Query Match 85.7%; Score 36; DB 2; Length 264;  
 Best Local Similarity 75.0%; Pred. No. 11;  
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGVLPQG 8  
 |||||:  
 DB 191 GGVLPQG 198

RESULT 10  
 F31485  
 Ig heavy chain V region (10-25) - mouse (fragment)

C:Species: Mus musculus (house mouse)  
 C:Date: 31-Jul-1989 #sequence\_revision 31-Jul-1989 #text\_change 30-May-1997  
 C:Accession: F31485  
 R:Bedzyk, W.D.; Johnson, L.S.; Riordan, G.S.; Voss Jr., E.W.  
 J. Biol. Chem. 264, 1565-1569, 1989  
 A:Title: Comparison of variable region primary structures within an anti-fluorescein  
 A:Reference number: A31485; MUID:89109167  
 A:Accession: F31485  
 A:Status: preliminary  
 A:Molecule type: protein  
 A:Residues: 1-32 <BED>  
 C:Superfamily: Immunoglobulin V region; immunoglobulin homology  
 C:Keywords: heterotetramer; immunoglobulin

Query Match 83.3%; Score 35; DB 2; Length 32;  
 Best Local Similarity 87.5%; Pred. No. 2.1;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGVLPQG 8  
 |||||:  
 DB 8 GGVLPQG 15

RESULT 11  
 D31485  
 Ig heavy chain V regions (3-24 and 5-14) - mouse (fragment)

C:Species: Mus musculus (house mouse)  
 C:Date: 31-Jul-1989 #sequence\_revision 31-Jul-1989 #text\_change 30-May-1997  
 C:Accession: D31485  
 R:Bedzyk, W.D.; Johnson, L.S.; Riordan, G.S.; Voss Jr., E.W.  
 J. Biol. Chem. 264, 1565-1569, 1989  
 A:Title: Comparison of variable region primary structures within an anti-fluorescein  
 A:Reference number: A31485; MUID:89109167  
 A:Accession: D31485  
 A:Status: preliminary  
 A:Molecule type: protein  
 A:Residues: 1-36 <BED>  
 C:Superfamily: Immunoglobulin V region; immunoglobulin homology  
 C:Keywords: heterotetramer; immunoglobulin

Query Match 83.3%; Score 35; DB 2; Length 36;  
 Best Local Similarity 87.5%; Pred. No. 2.4;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGVLPQG 8  
 |||||:  
 DB 8 GGVLPQG 15

RESULT 12  
 S33402  
 Ig heavy chain V region - mouse (fragment)

C:Species: Mus musculus (house mouse)  
 C:Date: 19-Mar-1997 #sequence\_revision 24-Jul-1997 #text\_change 17-Jul-1998  
 C:Accession: S33402; S36385  
 R:Kettlborough, C.A.; Saldanha, J.; Ansell, K.H.; Bendig, M.M.  
 Eur. J. Immunol. 23, 206-211, 1993  
 A:Title: Optimization of primers for cloning libraries of mouse immunoglobulin genes  
 A:Reference number: S33391; MUID:93122092

A:Accession: S33402  
 A:Molecule type: mRNA  
 A:Residues: 1-38 <KET>  
 A:Cross-references: EMBL:X73019  
 A:Experimental source: strain BALB/c  
 R:Ansell, K.H.  
 submitted to the EMBL Data Library, April 1993  
 A:Reference number: S36376  
 A:Accession: S36385  
 A:Molecule type: mRNA  
 A:Residues: 1-30 <ANS>  
 A:Cross-references: EMBL:X73019  
 C:Superfamily: Immunoglobulin V region; immunoglobulin homology  
 C:Keywords: heterotetramer; immunoglobulin  
 F:1-38/Domain: immunoglobulin homology (fragment) <IMM>

Query Match 83.3%; Score 35; DB 2; Length 38;  
 Best Local Similarity 87.5%; Pred. No. 2.5;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GGVLPQPG 8  
 |||||||  
 Db 6 GGGVQPG 13

## RESULT 13

A27606  
 Ig heavy chain V-a region (P26.9p2) - rabbit (fragments)  
 C:Species: Oryctolagus cuniculus (domestic rabbit)  
 C:Date: 08-Mar-1989 #sequence\_revision 30-Jun-1991 #text\_change 16-Aug-1996  
 C:Accession: A27606  
 R:Currier, S.J.; Gallarda, J.L.; Knight, K.L.  
 J. Immunol. 140, 1651-1659, 1988  
 A:Title: Partial molecular genetic map of the rabbit V-H chromosomal region.  
 A:Reference number: A27606; MUID:88154464  
 A:Accession: A27606  
 A:Molecule type: DNA  
 A:Residues: 1-59 <CUR>  
 A:Cross-references: GB:M19706  
 A:Note: The authors translated the codon TCC for residue 28 as Asp  
 C:Genetics:  
 A:Introns: 30/3  
 C:Superfamily: Immunoglobulin V region; immunoglobulin homology  
 C:Keywords: heterotetramer; immunoglobulin

Query Match 83.3%; Score 35; DB 2; Length 59;  
 Best Local Similarity 87.5%; Pred. No. 3.9;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GGVLPQPG 8  
 |||||||  
 Db 8 GGGVQPG 15

## RESULT 14

S36381  
 Ig heavy chain V region - mouse (fragment)  
 C:Species: Mus musculus (house mouse)  
 C:Date: 19-Mar-1997 #sequence\_revision 29-Aug-1997 #text\_change 23-Jul-1999  
 C:Accession: S36381; S33396  
 R:Ansell, K.H.  
 submitted to the EMBL Data Library, April 1993  
 A:Reference number: S36376  
 A:Accession: S36381  
 A:Molecule type: mRNA  
 A:Residues: 1-59 <ANS>  
 A:Cross-references: EMBL:X73012; NID:g295879; PIDN:CA51498.1; PID:g939941  
 R:Kettleborough, C.A.; Saldanha, J.; Ansell, K.H.; Bendig, M.M.  
 Eur. J. Immunol. 23, 206-211, 1993  
 A:Title: Optimization of primers for cloning libraries of mouse immunoglobulin genes  
 A:Reference number: S33391; MUID:93122092

A:Accession: S33396  
 A:Molecule type: mRNA  
 A:Residues: 3-43 <KET>  
 A:Cross-references: EMBL:X73012  
 A:Experimental source: strain BALB/c  
 C:Superfamily: immunoglobulin V region; immunoglobulin homology  
 C:Keywords: heterotetramer; immunoglobulin  
 F:14-59/Domain: immunoglobulin homology (fragment) <IMM>

Query Match 83.3%; Score 35; DB 2; Length 59;  
 Best Local Similarity 87.5%; Pred. No. 3.9;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GGVLPQPG 8  
 |||||||  
 Db 10 GGGVQPG 17

## RESULT 15

S36382  
 Ig heavy chain V region - mouse (fragment)  
 C:Species: Mus musculus (house mouse)  
 C:Date: 19-Mar-1997 #sequence\_revision 29-Aug-1997 #text\_change 23-Jul-1999  
 C:Accession: S36382; S33397  
 R:Ansell, K.H.  
 submitted to the EMBL Data Library, April 1993  
 A:Reference number: S36376  
 A:Accession: S36382  
 A:Molecule type: mRNA  
 A:Residues: 1-60 <ANS>  
 A:Cross-references: EMBL:X73011; NID:g295878; PIDN:CA51497.1; PID:g939940  
 R:Kettleborough, C.A.; Saldanha, J.; Ansell, K.H.; Bendig, M.M.  
 Eur. J. Immunol. 23, 206-211, 1993  
 A:Title: Optimization of primers for cloning libraries of mouse immunoglobulin genes  
 A:Reference number: S33391; MUID:93122092  
 A:Accession: S33397  
 A:Molecule type: mRNA  
 A:Residues: 1-33 <KET>  
 A:Cross-references: EMBL:X73011  
 A:Experimental source: strain BALB/c  
 C:Superfamily: immunoglobulin V region; immunoglobulin homology  
 C:Keywords: heterotetramer; immunoglobulin  
 F:8-60/Domain: immunoglobulin homology (fragment) <IMM>

Query Match 83.3%; Score 35; DB 2; Length 60;  
 Best Local Similarity 87.5%; Pred. No. 4;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GGVLPQPG 8  
 |||||||  
 Db 1 GGGVQPG 8

Search completed: June 13, 2001, 14:10:46  
 Job time: 147 sec



*This Page Blank (uspto)*

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 13, 2001, 14:21:46 ; Search time 74.44 Seconds

(without alignments)  
3.681 Million cell updates/sec

Title: PCT-US01-05825A-15

Perfect score: 42

Sequence: 1 GGVLVQPG 8

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SWISSPROT\_39:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	38	90.5	116	1	AMPN_SALTY
2	36	85.7	116	1	AMPN_SALTY
3	35	83.3	111	1	HY27_MOUSE
4	35	83.3	113	1	HY28_MOUSE
5	35	83.3	113	1	HY29_MOUSE
6	35	83.3	113	1	HY30_MOUSE
7	35	83.3	113	1	HY31_MOUSE
8	35	83.3	113	1	HY32_MOUSE
9	35	83.3	115	1	HY33_MOUSE
10	35	83.3	115	1	HY34_MOUSE
11	35	83.3	115	1	HY35_MOUSE
12	35	83.3	115	1	HY36_MOUSE
13	35	83.3	116	1	HY37_MOUSE
14	35	83.3	116	1	HY38_MOUSE
15	35	83.3	116	1	HY39_MOUSE
16	35	83.3	117	1	HY40_MOUSE
17	35	83.3	117	1	HY41_MOUSE
18	35	83.3	117	1	HY42_MOUSE
19	35	83.3	117	1	HY43_MOUSE
20	35	83.3	117	1	HY44_MOUSE
21	35	83.3	118	1	HY45_MOUSE
22	35	83.3	118	1	HY46_MOUSE
23	35	83.3	119	1	HY47_MOUSE
24	35	83.3	119	1	HY48_MOUSE
25	35	83.3	119	1	HY49_MOUSE
26	35	83.3	119	1	HY50_MOUSE
27	35	83.3	119	1	HY51_MOUSE
28	35	83.3	119	1	HY52_MOUSE
29	35	83.3	120	1	HY53_MOUSE
30	35	83.3	120	1	HY54_MOUSE
31	35	83.3	122	1	HY55_MOUSE
32	35	83.3	122	1	HY56_MOUSE
33	35	83.3	123	1	HY57_MOUSE

## ALIGNMENTS

34	35	83.3	123	1	HY19_MOUSE
35	35	83.3	123	1	HY22_MOUSE
36	35	83.3	123	1	HY23_MOUSE
37	35	83.3	123	1	HY24_MOUSE
38	35	83.3	123	1	HY25_MOUSE
39	35	83.3	136	1	HY16_MOUSE
40	35	83.3	142	1	HY01_RAT
41	35	83.3	144	1	HY26_MOUSE
42	35	83.3	238	1	Y457_CHLTR
43	34	81.0	115	1	HY3F_HUMAN
44	34	81.0	115	1	HY05_CARAV
45	33	78.6	238	1	Y573_CHLRN

RESULT 1	
HY30_HUMAN	STANDARD; PRT; 116 AA.
ID HY30_HUMAN	
AC P01778;	
DT 21-JUL-1986 (Rel. 01, Created)	
DR 21-JUL-1986 (Rel. 01, Last sequence update)	
DE 15-JUL-1999 (Rel. 38, Last annotation update)	
DE IG HEAVY CHAIN V-III REGION ZAP.	
OS Homo sapiens (Human).	
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;	
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
OX NCBI_TaxID=9606;	
RN [1]	
RP SEQUENCE.	
RX MEDLINE=74142702; PubMed=4522793;	
RA Capra J.D., Kenoe J.M.;	
RT "Variable region sequences of five human immunoglobulin heavy chains	
RT of the VH3 subgroup: definitive identification of four heavy chain	
RT hypervariable regions."	
RL Proc. Natl. Acad. Sci. U.S.A. 71:845-848(1974).	
CC -1- MISCELLANEOUS: THIS CHAIN WAS ISOLATED FROM AN IGAL MYELOMA	
CC PROTEIN.	
DR PIR: A02061; A1H02P.	
DR HSSP: P01772; 2IG2.	
DR InterPro: IPR003006; -	
DR Pfam: PF00047; Ig: 1.	
KW Immunoglobulin V region.	
FT NON_TER 116	
FT SEQUENCE 116 AA; 12582 MW; 892F8C217CEC9865 CRC64;	
Query Match	90.5%; Score 38; DB 1; Length 116;
Best Local Similarity	87.5%; Pred. No. 0.94;
Matches 7; Conservative	0; Mismatches 1; Indels 0; Gaps 0;
OY 1 GGVLVQPG 8	
DB 111111	
8 GGVLVQPG 15	
RESULT 2	
AMPN_SALTY	STANDARD; PRT; 263 AA.
ID AMPN_SALTY	
AC P10882;	
DT 01-JUL-1989 (Rel. 11, Created)	
DR 01-NOV-1991 (Rel. 20, Last sequence update)	
DE 30-MAY-2000 (Rel. 39, Last annotation update)	
DE METHIONINE AMINOPEPTIDASE (EC 3.4.11.18) (MAP) (PEPTIDASE M).	
GN MAP OR PEPM.	
OS Salmonella typhimurium.	
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;	
OC Salmonella.	
OX NCBI_TaxID=602;	
RN [1]	
RP SEQUENCE FROM N.A.	
RC STRAIN-LT2;	

```

OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=81013937; PubMed=6251474;
RA Bernard O., Gough N.M.;
RT "Nucleotide sequence of immunoglobulin heavy chain joining segments
RT between translocated VH and mu constant regions genes.";
RL Proc. Natl. Acad. Sci. U.S.A. 77:3630-3634(1980).
CC -1- MISCELLANEOUS: THE SEQUENCE OF THE FIRST 197 RESIDUES OF THE C
CC REGION WAS ALSO DETERMINED & DIFFERS IN ONLY 3 POSITIONS FROM THE
CC CORRESPONDING PORTION OF THE MOUSE MOPC 104E MU CHAIN.
DR PIR: A02074; MMS576.
DR InterPro: IPR003006; -.
DR Pfam: PF00047; Ig; 1.
KW Immunoglobulin V region.
FT NON_TER 1 1
FT 111 111
SQ SEQUENCE 111 AA; 12304 MW; 0EDE98EC7348056A CRC64;

Query Match
Best Local Similarity 83.3%; Score 35; DB 1; Length 111;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0

QY 1 GGVLPQPG 8
| | | | |
DB 3 GGGVLPQPG 10

RESULT 4
HV27_MOUSE STANDARD; PRT; 113 AA.
ID HV27_MOUSE
AC P01796;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 13-JUL-1999 (Rel. 38, Last annotation update)
DT IG HEAVY CHAIN V-III REGION A4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE.
RX MEDLINE=78158406; PubMed=417344;
RA Vrana M., Rudikoff S., Potter M.;
RT "Sequence variation among heavy chains from inulin-binding myeloma
RT proteins ";
RL Proc. Natl. Acad. Sci. U.S.A. 75:1957-1961(1978).
CC -1- MISCELLANEOUS: THIS CHAIN WAS ISOLATED FROM A MYELOMA PROTEIN THAT
CC BINDS INULIN.
DR PIR: A93818; AVMSAB.
DR InterPro: IPR003006; -.
DR Pfam: PF00047; Ig; 1.
KW Immunoglobulin V region.
FT DISULFID 22 98
FT NON_TER 113 113
FT 113 113
SQ SEQUENCE 113 AA; 12675 MW; 76658C12IC598285 CRC64;

Query Match
Best Local Similarity 83.3%; Score 35; DB 1; Length 113;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGVLPQPG 8
| | | | |
DB 8 GGGVLPQPG 15

RESULT 5
HV28_MOUSE

```



ID HV28\_MOUSE STANDARD; PRT; 113 AA.  
 AC P01797;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE IG HEAVY CHAIN V-III REGION U61.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NCBI\_TaxId=10090;  
 RN [1]  
 RP SEQUENCE.  
 RX MEDLINE=78158406; PubMed=417344;  
 RA Vrana M., Rudikoff S., Potter M.;  
 RT "Sequence variation among heavy chains from inulin-binding myeloma  
 proteins.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 75:1957-1961(1978).  
 CC -I- MISCELLANEOUS: THIS CHAIN WAS ISOLATED FROM A MYELOMA PROTEIN THAT  
 BINDS INULIN.  
 DR PIR: B93818; AVMS61.  
 DR InterPro: IPR003006; -.  
 DR Pfam: PF00047; 19; 1.  
 KW Immunoglobulin V region.  
 FT DISULFID 22 98  
 FT NON\_TER 113 113 BY SIMILARITY.  
 SO SEQUENCE 113 AA; 12671 MW; 76634C12C598285 CRC64;

Query Match 83.3%; Score 35; DB 1; Length 113;  
 Best Local Similarity 87.5%; Pred. No. 3.6;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GGVLPQPG 8  
 DB 8 GGVLPQPG 15

RESULT 6  
 ID HV29\_MOUSE STANDARD; PRT; 113 AA.  
 AC P01798;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE IG HEAVY CHAIN V-III REGION E109.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NCBI\_TaxId=10090;  
 RN [1]  
 RP SEQUENCE.  
 RX MEDLINE=78158406; PubMed=417344;  
 RA Vrana M., Rudikoff S., Potter M.;  
 RT "Sequence variation among heavy chains from inulin-binding myeloma  
 proteins.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 75:1957-1961(1978).  
 CC -I- MISCELLANEOUS: THIS CHAIN WAS ISOLATED FROM A MYELOMA PROTEIN THAT  
 BINDS INULIN.  
 DR PIR: C93818; AVMS09.  
 DR InterPro: IPR003006; -.  
 DR Pfam: PF00047; 19; 1.  
 KW Immunoglobulin V region.  
 FT DISULFID 22 98  
 FT NON\_TER 113 113 BY SIMILARITY.  
 SO SEQUENCE 113 AA; 12647 MW; EE50F2F20EDB129B CRC64;

Query Match 83.3%; Score 35; DB 1; Length 113;  
 Best Local Similarity 87.5%; Pred. No. 3.6;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GGVLPQPG 8  
 DB 8 GGVLPQPG 15

DB 8 GGVLPQPG 15  
 RESULT 7  
 ID HV30\_MOUSE STANDARD; PRT; 113 AA.  
 AC P01799;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE IG HEAVY CHAIN V-III REGION ABE-47N.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NCBI\_TaxId=10090;  
 RN [1]  
 RP SEQUENCE.  
 RX MEDLINE=77134726; PubMed=402936;  
 RA Vrana M., Rudikoff S., Potter M.;  
 RT "Heavy chain variable-region sequence from an inulin-binding myeloma  
 protein.";  
 RL Biochemistry 16:1170-1175(1977).  
 CC -I- MISCELLANEOUS: THIS CHAIN WAS ISOLATED FROM A MYELOMA PROTEIN THAT  
 BINDS INULIN.  
 DR PIR: A90400; AVMSB7.  
 DR InterPro: IPR003006; -.  
 DR Pfam: PF00047; 19; 1.  
 KW Immunoglobulin V region.  
 FT DISULFID 22 98  
 FT NON\_TER 113 113 BY SIMILARITY.  
 SO SEQUENCE 113 AA; 12675 MW; 76658C16C779845E CRC64;

Query Match 83.3%; Score 35; DB 1; Length 113;  
 Best Local Similarity 87.5%; Pred. No. 3.6;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GGVLPQPG 8  
 DB 8 GGVLPQPG 15

RESULT 8  
 ID HV31\_MOUSE STANDARD; PRT; 113 AA.  
 AC P01800;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 01-OCT-1993 (Rel. 27, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE IG HEAVY CHAIN V-III REGION T957.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NCBI\_TaxId=10090;  
 RN [1]  
 RP SEQUENCE.  
 RX MEDLINE=81216632; PubMed=6787122;  
 RA Rudikoff S., Potter M.;  
 RT "Immunoglobulin heavy chains from anti-inulin myeloma proteins:  
 evidence for a new heavy chain joining segment.";  
 RL J. Immunol. 127:191-194(1981).  
 CC -I- MISCELLANEOUS: THIS CHAIN WAS ISOLATED FROM A MYELOMA PROTEIN THAT  
 BINDS INULIN.  
 DR PIR: A92810; AVMS57.  
 DR InterPro: IPR003006; -.  
 DR Pfam: PF00047; 19; 1.  
 KW Immunoglobulin V region.  
 FT DISULFID 22 98  
 FT NON\_TER 113 113 BY SIMILARITY.  
 SO SEQUENCE 113 AA; 12732 MW; 26618F626B59859E CRC64;

Query Match 83.3%; Score 35; DB 1; Length 113;

Best Local Similarity 87.5%; Pred. No. 3.6;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GGVLPQPG 8  
111111  
Db 8 GGGVLPQPG 15

## RESULT 9

HV34\_MOUSE  
ID HV34\_MOUSE STANDARD; PRT; 113 AA.  
AC P01803;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 21-JUL-1986 (Rel. 01, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE IG HEAVY CHAIN V REGION AMPC1.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_Taxid=10090;  
RN [1]  
RP SEQUENCE.  
RX MEDLINE-81216632; PubMed-6787122;  
RA Rudikoff S., Potter M.;  
RT "Immunoglobulin heavy chains from anti-inulin myeloma proteins:  
RL evidence for a new heavy chain joining segment.";  
J. Immunol. 127:191-194(1981).  
CC -1- MISCELLANEOUS: THIS CHAIN WAS ISOLATED FROM A MYELOMA PROTEIN THAT  
CC BINDS INULIN.  
DR PIR: A02073; HYMSAM.  
DR InterPro: IPR003006; .  
DR Pfam: PF00047; 1g; 1.  
KW Immunoglobulin V region.  
FT DISULFID 22 98 BY SIMILARITY.  
FT NON\_TER 113 113  
SQ SEQUENCE 113 AA; 12691 MW; 7A6D906AA966E9E CRC64;

Query Match 83.3%; Score 35; DB 1; Length 113;  
Best Local Similarity 87.5%; Pred. No. 3.6;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GGVLPQPG 8  
111111  
Db 8 GGGVLPQPG 15

## RESULT 10

HV32\_MOUSE  
ID HV32\_MOUSE STANDARD; PRT; 115 AA.  
AC P01801;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 21-JUL-1986 (Rel. 01, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE IG HEAVY CHAIN V-III REGION J606.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_Taxid=10090;  
RN [1]  
RP SEQUENCE.  
RX MEDLINE-82099361; PubMed-6798111;  
RA Johnson N., Stankard J., Paul L., Hood L.;  
RT "The complete V domain amino acid sequences of two myeloma inulin-  
binding proteins.";  
J. Immunol. 128:302-307(1982).  
CC -1- MISCELLANEOUS: THIS CHAIN WAS ISOLATED FROM A MYELOMA PROTEIN THAT  
CC BINDS INULIN.  
DR PIR: C92811; AVMS06.  
DR InterPro: IPR003006; .  
DR Pfam: PF00047; 1g; 1.  
KW Immunoglobulin V region.  
FT DISULFID 22 98 BY SIMILARITY.  
FT NON\_TER 115 115

FT NON\_TER 115 115  
SQ SEQUENCE 115 AA; 12810 MW; B67AD6638A121A5F CRC64;

Query Match 83.3%; Score 35; DB 1; Length 115;  
Best Local Similarity 87.5%; Pred. No. 3.6;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GGVLPQPG 8  
111111  
Db 8 GGGVLPQPG 15

## RESULT 11

HV33\_MOUSE  
ID HV33\_MOUSE STANDARD; PRT; 115 AA.  
AC P01802;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 21-JUL-1986 (Rel. 01, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE IG HEAVY CHAIN V-III REGION W3082.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_Taxid=10090;  
RN [1]  
RP SEQUENCE.  
RX MEDLINE-82099361; PubMed-6798111;  
RA Johnson N., Stankard J., Paul L., Hood L.;  
RT "The complete V domain amino acid sequences of two myeloma inulin-  
binding proteins.";  
J. Immunol. 128:302-307(1982).  
CC -1- MISCELLANEOUS: THIS CHAIN WAS ISOLATED FROM A MYELOMA PROTEIN THAT  
CC BINDS INULIN.  
DR PIR: D92811; AVMS82; .  
DR InterPro: IPR003006; .  
DR Pfam: PF00047; 1g; 1.  
KW Immunoglobulin V region.  
FT DISULFID 22 98 BY SIMILARITY.  
FT NON\_TER 115 115  
SQ SEQUENCE 115 AA; 12887 MW; 9B4517648C121C5A CRC64;

Query Match 83.3%; Score 35; DB 1; Length 115;  
Best Local Similarity 87.5%; Pred. No. 3.6;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GGVLPQPG 8  
111111  
Db 8 GGGVLPQPG 15

## RESULT 12

HV3D\_HUMAN  
ID HV3D\_HUMAN STANDARD; PRT; 115 AA.  
AC P01765;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 21-JUL-1986 (Rel. 01, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE IG HEAVY CHAIN V-III REGION T1L.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_Taxid=9606;  
RN [1]  
RP SEQUENCE.  
RX MEDLINE-76005528; PubMed-409716;  
RA Wang A.-C., Wang I.-Y., Fudenberg H.H.;  
RT "Immunoglobulin structure and genetics. Identity between variable  
regions of a mu and a gamma2 chain.";  
J. Biol. Chem. 252:7192-7199(1977).  
CC -1- MISCELLANEOUS: THE SEQUENCES OF THE V REGIONS OF THE HEAVY CHAINS  
OF IGM AND IGG2 ISOLATED FROM A SINGLE PATIENT WITH BICLONAL

CC GAMMOPATHY ARE IDENTICAL. THEIR LIGHT CHAINS ARE APPARENTLY ALSO  
 CC IDENTICAL.  
 DR PIR: A02048; H3HUTL.  
 DR InterPro: IPR003006; -  
 DR Pfam: PF00047; 19; 1.  
 KW Immunoglobulin V region.  
 FT NON\_TER 115 115  
 SO SEQUENCE 115 AA; 12356 MW; 4DC67D179F62326 CRC64;

Query Match 83.3%; Score 35; DB 1; Length 115;  
 Best Local Similarity 87.5%; Pred. No. 3.6;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GCVLVQPG 8  
 111111  
 DB 8 GCGLVQPG 15

RESULT 13  
 HV36\_MOUSE STANDARD; PRT; 116 AA.  
 ID HV36\_MOUSE  
 AC P01806;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE IG HEAVY CHAIN V REGION 441 PRECURSOR.  
 OS Mus musculus (Mouse).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
 NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=62058449; PubMed=6795591;  
 RA O'illo R., Aufray C., Sikorav J.-L., Rougeon F.;  
 RT "Mouse heavy chain variable regions: nucleotide sequence of a  
 germ-line VH gene segment."  
 RL Nucleic Acids Res. 9:4099-4109(1981).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
 DR EMBL: V00774; CAA24152.1; -  
 DR PIR: A02076; HVMS44.  
 DR HSSP: P01810; 2FBJ.  
 DR InterPro: IPR003006; -  
 DR Pfam: PF00047; 19; 1.  
 KW Immunoglobulin V region; Signal.  
 FT SIGNAL 1 18  
 FT CHAIN 19 116 IG HEAVY CHAIN V REGION 441.  
 FT NON\_TER 116 116  
 SO SEQUENCE 116 AA; 12911 MW; 6B4FF3EF9A95D9F3 CRC64;

Query Match 83.3%; Score 35; DB 1; Length 116;  
 Best Local Similarity 87.5%; Pred. No. 3.7;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GCVLVQPG 8  
 111111  
 DB 26 GCGLVQPG 33

RESULT 14  
 HV3R\_HUMAN STANDARD; PRT; 116 AA.  
 ID HV3R\_HUMAN  
 AC P01779;  
 DT 21-JUL-1986 (Rel. 01, Created)

DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE IG HEAVY CHAIN V-II REGION TGR.  
 OS Homo sapiens (Human).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE.  
 RX MEDLINE=74142702; PubMed=4522793;  
 RA Capra J.D., Kehoe J.M.;  
 RT "Variable region sequences of five human immunoglobulin heavy chains  
 RT of the VH3 subgroup; definitive identification of four heavy chain  
 RT hypervariable regions."  
 RL Proc. Natl. Acad. Sci. U.S.A. 71:845-848(1974).  
 CC -1- MISCELLANEOUS: THIS CHAIN WAS ISOLATED FROM AN IGAI MYELOMA  
 CC PROTEIN.  
 DR PIR: A02062; A1HUTU.  
 DR InterPro: IPR003006; -  
 DR Pfam: PF00047; 19; 1.  
 KW Immunoglobulin V region.  
 FT NON\_TER 116 116  
 SO SEQUENCE 116 AA; 12431 MW; EB705F53A963F0C CRC64;

Query Match 83.3%; Score 35; DB 1; Length 116;  
 Best Local Similarity 87.5%; Pred. No. 3.7;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GCVLVQPG 8  
 111111  
 DB 8 GCGLVQPG 15

RESULT 15  
 HV3T\_HUMAN STANDARD; PRT; 116 AA.  
 ID HV3T\_HUMAN  
 AC P01781;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE IG HEAVY CHAIN V-II REGION GAL.  
 OS Homo sapiens (Human).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE.  
 RX MEDLINE=75059123; PubMed=4803843;  
 RA Watanabe S., Barnikol H.U., Horn J., Bettram J., Hilschmann N.;  
 RT "The primary structure of a monoclonal IgM-immunoglobulin  
 RT (macroglobulin Gal.), II: the amino acid sequence of the H-chain (mu-  
 RT type), subgroup H III. Architecture of the complete IgM-molecule."  
 RL Hoppe-seyler's Z. Physiol. Chem. 354:1505-1509(1973).

CC REVISION TO THE COMPOSITION OF 28-33.  
 RA Hilschmann N.;  
 RL Submitted (JUN-1975) to the PIR data bank.  
 CC -1- MISCELLANEOUS: THIS MU CHAIN WAS ISOLATED FROM A WALDENSTROM'S  
 CC MACROGLOBULIN.  
 DR PIR: A02064; M3HUGL.  
 DR HSSP: P01772; 21G2.  
 DR InterPro: IPR003006; -  
 DR Pfam: PF00047; 19; 1.  
 KW Immunoglobulin V region.  
 FT NON\_TER 116 116  
 SO SEQUENCE 116 AA; 12730 MW; 2C67CA9AAAAA1282 CRC64;

Query Match 83.3%; Score 35; DB 1; Length 116;  
 Best Local Similarity 87.5%; Pred. No. 3.7;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Wed Jun 13 14:59:49 2001

pct-us01-05825a-15.rsp

Page 6

Oy 1 GGVLVQPG 8  
|| |||||  
Db 8 GGDVQPG 15

Search completed: June 13, 2001, 14:21:46  
Job time: 806 sec

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 13, 2001, 14:20:34 ; Search time 225.85 seconds  
(without alignments)  
4.152 Million cell updates/sec

Title: PCT-US01-05825a-15

Perfect score: 42

Sequence: 1 GGVLVQPG 8

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 374700 seqs, 117207915 residues

Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :  
1: SP. archaea: \*  
2: SP. bacteria: \*  
3: SP. fungi: \*  
4: SP. human: \*  
5: SP. invertebrate: \*  
6: SP. mammal: \*  
7: SP. mhc: \*  
8: SP. organelle: \*  
9: SP. phage: \*  
10: SP. plant: \*  
11: SP. rodent: \*  
12: SP. unclassified: \*  
13: SP. vertebrate: \*  
14: SP. virus: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	38	90.5	510	2	068278 cytophaga x
2	36	85.7	251	2	09RS49 streptomyces
3	36	85.7	253	2	09XA16 streptomyces
4	35	83.3	15	11	090V16 ratulus sp.
5	35	83.3	95	4	09UL86 homo sapien
6	35	83.3	118	4	09UL91 homo sapien
7	35	83.3	118	4	09UL72 homo sapien
8	35	83.3	238	2	09PY75 chlamydia m
9	35	83.3	257	1	057998 pyrococcus
10	35	83.3	272	1	09V2H0 pyrococcus
11	35	83.3	323	13	073907 gallus gall
12	35	83.3	563	2	052169 pseudomonas
13	35	83.3	563	2	006573 pseudomonas
14	35	83.3	1144	2	09UX22 neisseria m
15	35	83.3	1144	2	09UX22 neisseria m
16	34	81.0	218	2	09VX8 micrococcus
17	34	81.0	399	5	09NGF2 drosophila
18	34	81.0	399	5	09NGA6 drosophila
19	34	81.0	444	5	010923 caenorhabdit

20	34	81.0	812	5	006452 ephydria m
21	34	81.0	1241	10	09LHK4 arabidopsis
22	34	81.0	1342	10	09ZS84 lycopersico
23	34	81.0	2189	5	09VXD5 arabidopsis
24	33	78.6	190	10	09STW2 chlamydia p
25	33	78.6	238	2	09J012 chlamydia p
26	33	78.6	260	5	09VEH2 drosophila
27	33	78.6	325	10	043751 capsicum an
28	33	78.6	485	10	096547 capsicum an
29	33	78.6	489	10	042872 lycopersico
30	33	78.6	926	10	09LZS4 arabidopsis
31	33	78.6	1207	13	09PVE5 brachydanio
32	33	76.2	113	4	09UL90 homo sapien
33	33	76.2	116	4	09UL93 homo sapien
34	33	76.2	121	4	09UL71 homo sapien
35	33	76.2	122	4	09UL84 homo sapien
36	32	76.2	147	4	09Y509 homo sapien
37	32	76.2	168	2	09S2V7 streptomyces
38	32	76.2	238	10	09ZNS9 oryza sativ
39	32	76.2	253	2	09PEW8 xylella fas
40	32	76.2	306	14	09WQ76 turkey ente
41	32	76.2	316	10	09ZQ25 arabidopsis
42	32	76.2	330	2	054110 streptomyces
43	32	76.2	358	10	09LPI5 arabidopsis
44	32	76.2	363	1	09YCS0 aetopyrum p
45	32	76.2	610	2	086712 streptomyces

## ALIGNMENTS

RESULT	1	PRELIMINARY:	PRT:	510 AA.
ID	068278	068278		
AC	068278			
DT	01-AUG-1998 (TREMUREL. 07, Created)			
DT	01-AUG-1998 (TREMUREL. 07, Last sequence update)			
DT	01-NOV-1998 (TREMUREL. 08, Last annotation update)			
DE	ARFI (EC 3.2.1.55) (ALPHA-L-ARABINOFURANOSIDASE) (ARABINOSIDASE).			
GN	ARFI.			
OS	Cytophaga xylanolytica.			
OC	Bacteria; CFB group; Cytophagales; Cytophagaceae; Cytophaga.			
OX	NCBI_TaxID:990;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-XM3;			
RX	MEDLINE-98247330; PubMed-9572972;			
RA	Kim K.S., Lilburn T.G., Renner M.J., Breznak J.A.;			
RT	"Arfi and arfi, two genes encoding alpha-L-arabinofuranosidases in			
RL	Cytophaga xylanolytica";			
CC	Appl. Environ. Microbiol. 64:1919-1923(1998).			
CC	"- CATALYTIC ACTIVITY: HYDROLYSIS OF TERMINAL NON-REDUCING ALPHA-L-			
DR	ARABINOFURANOSIDE RESIDUES IN ALPHA-L-ARABINOSIDES.			
KW	EMBL: AF028018; AAC38456.1; "			
SO	SEQUENCE 510 AA; 57571 MW; BCE638E4DBC2A074 CRC64;			
QY	1 GGVLVQPG 8			
DB	54 GGVLVQPG 61			
Query Match	90.5%; Score 38; DB 2; Length 510;			
Best Local Similarity	75.0%; Pred. No. 24;			
Matches	6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;			
RESULT	2	PRELIMINARY:	PRT:	251 AA.
ID	09RS49			
AC	09RS49			
DT	01-MAY-2000 (TREMUREL. 13, Created)			
DT	01-MAY-2000 (TREMUREL. 13, Last sequence update)			

DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)  
 DE HYPOTHETICAL 26.8 KDA PROTEIN.  
 GN DR2216.  
 OS Deinococcus radiodurans.  
 OC Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.  
 OX NCBI\_TaxID=1299;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=RL;  
 RX MEDLINE=20036896; PubMed=10567266;  
 RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,  
 RA Dodson R.J., Helt D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,  
 RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,  
 RA Vamathevan J.J., Lam P., McDonald L., Utterback T., Zaleski C.,  
 RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,  
 RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,  
 RA Fraser C.M.;  
 RT "Genome Sequence of the Radioresistant Bacterium Deinococcus  
 RT radiodurans R1.";  
 RL Science 266:1571-1577(1999).  
 DR EMBL; AE002054; AAF11756.1; -;  
 DR TIGR; DR2216; -;  
 KW Hypothetical protein.  
 SO SEQUENCE 251 AA; 26788 MW; 3DCA855F26D5A847 CRC64;

Query Match 85.7%; Score 36; DB 2; Length 251;  
 Best Local Similarity 75.0%; Pred. No. 27;  
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 GGVLPQPG 8  
 |||||  
 Db 51 GGVLPQPG 58

RESULT 3  
 ID 09XAL6 PRELIMINARY; PRT; 263 AA.  
 AC 09XAL6;  
 DT 01-NOV-1999 (TREMBLrel. 12, Created)  
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)  
 DE PUTATIVE HYDROXYLASE.  
 GN SC4C6.24C.  
 OS Streptomyces coelicolor.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OX NCBI\_TaxID=1902;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A3(2);  
 RA Seeger S., Harris D.;  
 RT "A set of ordered cosmids and a detailed genetic and physical map for  
 RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";  
 RL Submitted (JUN-1999) to the EMBL/Genbank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A3(2);  
 RA James K.D., Parkhill J., Barrell B.G., Rajandream M.A.,  
 RL Submitted (JUN-1999) to the EMBL/Genbank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A3(2);  
 RX MEDLINE=97000351; PubMed=8843436;  
 RA Redenbach M., Kleser H.W., Denaplatte D., Elchner A., Cullum J.,  
 RA Kinashi H., Hopwood D.A.;  
 RT "A set of ordered cosmids and a detailed genetic and physical map for  
 RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";  
 RL MBL; AL079355; CAB45583.1; -;  
 DR EMBL; AL079355; CAB45583.1; -;  
 SO SEQUENCE 263 AA; 27276 MW; F9A9F43F8E4976F9 CRC64;

Query Match 85.7%; Score 36; DB 2; Length 263;  
 Best Local Similarity 100.0%; Pred. No. 29;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GGVLPQPG 7  
 |||||  
 Db 92 GGVLPQPG 98

RESULT 4  
 ID 09QV16 PRELIMINARY; PRT; 15 AA.  
 AC 09QV16;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DE PROLACTIN-BINDING PROTEIN (FRAGMENT).  
 OS Rattus sp.  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.  
 OX NCBI\_TaxID=10118;  
 RN [1]  
 RP SEQUENCE.  
 RX MEDLINE=95094032; PubMed=8000909;  
 RA Cohen H., Cohen O., Gagnon J.;  
 RT "Serum prolactin-binding protein (PRL-BP) of human and rat are  
 RT identified as igc.";  
 RL C.R. Acad. Sci. III, Sci. Vie 317:293-298(1994).  
 DR HSSP; P01789; IWCP.  
 SO SEQUENCE 15 AA; 1469 MW; 35ED2512FF3FA369 CRC64;

Query Match 83.3%; Score 35; DB 11; Length 15;  
 Best Local Similarity 87.5%; Pred. No. 2;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GGVLPQPG 8  
 |||||  
 Db 8 GGVLPQPG 15

RESULT 5  
 ID 09ULB6 PRELIMINARY; PRT; 95 AA.  
 AC 09ULB6;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DE IMUNOGLOBULIN HEAVY CHAIN (FRAGMENT).  
 GN VH.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Tange Y., Kayano H.;  
 RT "Human VH gene sequence.";  
 RL Submitted (NOV-1999) to the EMBL/Genbank/DBJ databases.  
 DR EMBL; AB035268; BAA87067.1; -;  
 DR HSSP; P01772; ZFBA.  
 DR INTERPRO; IPR003006; -;  
 DR PRAM; PF00047; 1g; 1.  
 FT NON\_TER 1 1  
 FT NON\_TER 95 95  
 SO SEQUENCE 95 AA; 10527 MW; 90A8C6D16D22574A CRC64;

Query Match 83.3%; Score 35; DB 4; Length 95;  
 Best Local Similarity 87.5%; Pred. No. 15;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GGVLPQPG 8

Db 7 GGLVQPG 14

RESULT 6  
09UL91 PRELIMINARY; PRT; 118 AA.

AC 09UL91;  
DT 01-MAY-2000 (TREMBLrel. 13, Created)  
DT 01-JUN-2000 (TREMBLrel. 14, Last sequence update)  
DE MYOSIN-REACTIVE IMMUNOGLOBULIN HEAVY CHAIN VARIABLE REGION (FRAGMENT).  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=98277139; PubMed=9614934;  
RA Wu X., Liu B., Van der Merwe P.L., Kalis N.N., Berney S.M.,  
RA Young D.C.;  
RT "Myosin-reactive autoantibodies in rheumatic carditis and normal  
RT fetus";  
RL Clin. Immunol. Immunopathol. 87:184-192(1998).  
DR EMBL; AF035023; AAD56259.1; -  
DR INTERPRO: IPR003006; -  
DR PFAM; PF00047; 19; 1.  
FT NON\_TER 1  
FT NON\_TER 118  
SQ SEQUENCE 118 AA; 12843 MW; D0633949F2AC149D CRC64;

Query Match 83.3%; Score 35; DB 4; Length 118;  
Best Local Similarity 87.5%; Pred. No. 19;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 GGLVQPG 8  
Db 8 GGLVQPG 15

RESULT 7  
09UL72 PRELIMINARY; PRT; 118 AA.  
AC 09UL72;  
DT 01-MAY-2000 (TREMBLrel. 13, Created)  
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)  
DE MYOSIN-REACTIVE IMMUNOGLOBULIN HEAVY CHAIN VARIABLE REGION (FRAGMENT).  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=98277139; PubMed=9614934;  
RA Wu X., Liu B., Van der Merwe P.L., Kalis N.N., Berney S.M.,  
RA Young D.C.;  
RT "Myosin-reactive autoantibodies in rheumatic carditis and normal  
RT fetus";  
RL Clin. Immunol. Immunopathol. 87:184-192(1998).  
DR EMBL; AF035042; AAD56278.1; -  
DR HSSP; P01772; 2FB4.  
DR INTERPRO: IPR003006; -  
DR PFAM; PF00047; 19; 1.  
FT NON\_TER 1  
FT NON\_TER 118  
SQ SEQUENCE 118 AA; 12872 MW; BAD1A5944B2D5CCA CRC64;

Query Match 83.3%; Score 35; DB 4; Length 118;  
Best Local Similarity 87.5%; Pred. No. 19;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 GGLVQPG 8  
Db 8 GGLVQPG 15

RESULT 8  
09PJT5 PRELIMINARY; PRT; 238 AA.

AC 09PJT5;  
DT 01-OCT-2000 (TREMBLrel. 15, Created)  
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
DE HYPOTHETICAL PROTEIN TC0742.  
GN TC0742.  
OS Chlamydia muridarum.  
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
OX NCBI\_TaxID=83560;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MORN / NIGG;  
RX MEDLINE=20150255; PubMed=10684935;  
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,  
RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,  
RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,  
RA Gwynn M., Nelson W., Deboy R., Kolonay J., McClarty G., Salzberg S.L.,  
RA Eisen J., Fraser C.M.;  
RT "Genome sequences of Chlamydia trachomatis MORN and Chlamydia  
RT pneumoniae AR39.";  
RL Nucleic Acids Res. 28:1397-1406(2000).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Read T.D., Brunham R., Shen C., Gill S.R., Heidelberg J.F., White O.,  
RA Hickey E.K., Peterson J., Umeyam L.A., Utterback T., Berry K.,  
RA Bass S., Linher K., Weidman J., Khouri H., Craven B., Bowman C.,  
RA Dodson R., Gwynn M., Nelson W., Deboy R., Kolonay J., McClarty G.,  
RA Salzberg S.L., Eisen J., Fraser C.M.;  
RL Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.  
DR EMBL; AE002342; AAF39551.1; -  
DR TIGR; TC0742; -  
DR INTERPRO: IPR002876; -  
DR PFAM; PF01709; DUF28; 1.  
KW Hypothetical protein.  
SQ SEQUENCE 238 AA; 26486 MW; C5D5657B0909264C CRC64;

Query Match 83.3%; Score 35; DB 2; Length 238;  
Best Local Similarity 75.0%; Pred. No. 40;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Oy 1 GGLVQPG 8  
Db 120 GGLVQPG 127

RESULT 9  
057998 PRELIMINARY; PRT; 257 AA.  
AC 057998;  
DT 01-AUG-1998 (TREMBLrel. 07, Created)  
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)  
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)  
DE HYPOTHETICAL 28.7 KDA PROTEIN PH0260.  
GN PH0260.  
OS Pyrococcus horikoshii.  
OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.  
OX NCBI\_TaxID=53953;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=OT3;  
RX MEDLINE=98444137; PubMed=9679194;  
RA Kawarayashi Y., Savada M., Horikawa H., Halkawa Y., Hino Y.,  
RA Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,

RA Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y.,  
 RA Funahashi T., Tanaka T., Kudo Y., Yamazaki J., Kushiida N., Oguchi A.,  
 RA Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,  
 RA Masuchi Y., Shizuya H., Kikuchi H.,  
 RT "Complete sequence and gene organization of the genome of a hyper-  
 thermophilic archaeobacterium, Pyrococcus horikoshii OT3."  
 RL DNA Res. 5:55-76(1998).  
 DR EMBL: AP000001; BAA29332.1; -  
 DR INTERPRO: IPR001279; -  
 DR PFAM: PF00753; Lactamase\_B; 1.  
 KW Hypothetical protein  
 SQ SEQUENCE 257 AA; 28680 MW; C9249670AB73FA7E CRC64;

Query Match 83.3%; Score 35; DB 1; Length 257;  
 Best Local Similarity 62.5%; Pred. No. 44;  
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 GGVLP0PG 8  
 DB 44 GGILDPG 51

RESULT 10  
 ID 09V2H0 PRELIMINARY; PRT; 272 AA.  
 AC 09V2H0;  
 DT 01-MAY-2000 (TREMblrel. 13, Created)  
 DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)  
 DT 01-JUN-2000 (TREMblrel. 14, Last annotation update)  
 DE HYPOTHEICAL 30.4 KDA PROTEIN.  
 GN PAB2280.  
 OS Pyrococcus abyssi.  
 OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.  
 OX NCBI\_TaxID=29292;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-ORSAV;  
 RA Helling R.;  
 RT "Pyrococcus abyssi genome sequence: Insights into archaeal chromosome  
 structure and evolution."  
 RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AJ248283; CAB49028.1; -  
 KW Hypothetical protein.  
 SQ SEQUENCE 272 AA; 30372 MW; AE0250A0399E01AA CRC64;

Query Match 83.3%; Score 35; DB 1; Length 272;  
 Best Local Similarity 75.0%; Pred. No. 47;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 GGVLP0PG 8  
 DB 179 GGILVTPG 186

RESULT 11  
 ID 073907 PRELIMINARY; PRT; 323 AA.  
 AC 073907;  
 DT 01-AUG-1998 (TREMblrel. 07, Created)  
 DT 01-AUG-1998 (TREMblrel. 07, Last sequence update)  
 DT 01-MAY-2000 (TREMblrel. 13, Last annotation update)  
 DE CNKX-2.3 PROTEIN.  
 GN CNKX-2.3.  
 OS Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 OC Gallus.  
 OX NCBI\_TaxID=9031;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=HEART;

RX MEDLINE-96391182; PubMed-8798155;  
 RA Buchberger A., Pabst O., Brandt T., Seidl K., Arnold H.H.;  
 RT "Chick CNKX-2.3 represents a novel family member of vertebrate  
 RT homologues to the Drosophila homeobox gene tinman: differential  
 RT expression of CNKX-2.3 and CNKX-2.5 during heart and gut  
 development."  
 RL Mech. Dev. 56:151-163(1996).  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).  
 DR EMBL: X97667; CAA66257.1; -  
 DR HSSP: P23441; 1FTF.  
 DR INTERPRO: IPR001356; -  
 DR PFAM: PF00046; homeobox; 1.  
 DR PRINTS: PR00024; HOMEBOX.  
 DR PROSITE: PS00027; HOMEBOX\_1; 1.  
 DR PROSITE: PS50071; HOMEBOX\_2; 1.  
 KW Homeobox; DNA-binding; Nuclear protein.  
 SQ SEQUENCE 323 AA; 36034 MW; 09E644D7D54470CB CRC64;

Query Match 83.3%; Score 35; DB 13; Length 323;  
 Best Local Similarity 75.0%; Pred. No. 56;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 GGVLP0PG 8  
 DB 308 GGALQPG 315

RESULT 12  
 ID 052169 PRELIMINARY; PRT; 563 AA.  
 AC 052169;  
 DT 01-NOV-1996 (TREMblrel. 01, Created)  
 DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)  
 DT 01-OCT-2000 (TREMblrel. 15, Last annotation update)  
 DE PHER PROTEIN.  
 GN PHER.  
 OS Pseudomonas putida.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
 OC Pseudomonas.  
 OX NCBI\_TaxID=303;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BH;  
 RA Takeo M.;  
 RT "Cloning and sequencing of the regulation gene of the phenol  
 RT degradative genes from Pseudomonas putida BH."  
 RL Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: D63814; BAA09883.1; -  
 DR HSSP: P11028; 1FTF.  
 DR INTERPRO: IPR002078; -  
 DR PFAM: PF00158; sigma54; 1.  
 DR PROSITE: PS00675; SIGMA54\_INTERACT\_1; 1.  
 DR PROSITE: PS00676; SIGMA54\_INTERACT\_2; 1.  
 DR PROSITE: PS00688; SIGMA54\_INTERACT\_3; 1.  
 DR PROSITE: PS50045; SIGMA54\_INTERACT\_4; 1.  
 SQ SEQUENCE 563 AA; 63396 MW; 3CCD11001786F281 CRC64;

Query Match 83.3%; Score 35; DB 2; Length 563;  
 Best Local Similarity 85.7%; Pred. No. 1e+02;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 GGVLP0PG 8  
 DB 496 GGILQPG 502

RESULT 13  
 ID 006573 PRELIMINARY; PRT; 563 AA.  
 AC 006573;  
 AC 006573;



```

DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DT 01-OCT-2000 (TREMblrel. 15, Last annotation update)
DE TRANSCRIPTIONAL REGULATORY PROTEIN DMPR.
GN DMPR.
OS Pseudomonas sp. CF600.
OG Plasmid pVII50.
OC Bacteria; Proteobacteria.
OX NCBI_TaxID=306;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93194783; PubMed=8449869;
RA Shingler V., Bartilson M., Moore J.T.;
RT "Cloning and nucleotide sequence of the gene encoding the positive
RT regulator (Dmpr) of the phenol catabolic pathway encoded by pVII50 and
RT identification of Dmpr as a member of the NtrC family of
RT transcriptional activators."
RL J. Bacteriol. 175:1596-1604(1993).
CC -1- FUNCTION: REGULATORY PROTEIN OF THE PHENOL CATABOLIC PATHWAY.
CC -1- SIMILARITY: THE CENTRAL REGION CONTAINS A SIGMA-34 FACTOR
CC INTERACTION ATP-BINDING DOMAIN.
CC -1- SIMILARITY: BELONGS TO THE NTRC FAMILY OF TRANSCRIPTIONAL
CC ACTIVATORS.
CC -1- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-37 IS THE INITIATOR.
DR EMBL: X68033; CAA48174.1; -.
DR HSP; P11028; ITP.
DR INTERPRO: IPR002078; -.
DR INTERPRO: IPR002197; -.
DR PFAM: PF00158; sigma54.1.
DR PROSITE: PS00675; SIGMA54_INTERACT_1; 1.
DR PROSITE: PS00676; SIGMA54_INTERACT_2; 1.
DR PROSITE: PS00688; SIGMA54_INTERACT_3; 1.
DR PROSITE: PS50045; SIGMA54_INTERACT_4; 1.
KW Transcription regulation; DNA-binding; Activator; Plasmid;
KW ATP-binding.
FT DOMAIN 230 465 SIGMA-54 FACTOR INTERACTION (POTENTIAL).
FT NP_BIND 263 270 ATP (BY SIMILARITY).
FT NP_BIND 326 335 ATP (BY SIMILARITY).
FT DN_BIND 518 538 H-T-H MOTIF (BY SIMILARITY).
SQ SEQUENCE 563 AA; 63197 MW; 9261B4EE3A2F98BD CRC64;

Query Match
Best Local Similarity 83.3%; Score 35; DB 2; Length 563;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 GVLVOPG 8
DB 496 GVLVOPG 502

RESULT 14
Q9JXZ2 PRELIMINARY; PRT; 1144 AA.
AC Q9JXZ2;
DT 01-OCT-2000 (TREMblrel. 15, Created)
DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
DT 01-OCT-2000 (TREMblrel. 15, Last annotation update)
DE DNA POLYMERASE III, ALPHA SUBUNIT.
GN NMB1827.
OS Neisseria meningitidis (serogroup B).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=491;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=MC58 / SEROGROUP B;
RX MEDLINE=20175755; PubMed=10710307;
RA Tetschlin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
RA Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,
RA Nelson W.C., Gwin M.L., DeBoy R., Peterson J.D., Hickey E.K.,
RA Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
RA Mason T., Clecio A., Parksey D.S., Blair E., Cifton H., Clark E.B.,
RA Colton M.D., Uterback T.R., Khouri H., Qin H., Vamathavan J.,

```

```

RA Gill J., Scarlato V., Maignani V., Pizzo M., Grandi G., Sun L.,
RA Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;
RT "Complete genome sequence of Neisseria meningitidis serogroup B strain
RT MC58."
RL Science 287:1809-1815(2000).
DR EMBL: AE002532; AAF42162.1; -.
DR TIGR: NMB1827; -.
SQ SEQUENCE 1144 AA; 127134 MW; 4CEAB86F90DD7EA CRC64;

Query Match
Best Local Similarity 83.3%; Score 35; DB 2; Length 1144;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGVLPVG 8
DB 518 GGVLPVG 525

RESULT 15
Q9JYX8 PRELIMINARY; PRT; 1144 AA.
AC Q9JYX8;
DT 01-OCT-2000 (TREMblrel. 15, Created)
DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
DT 01-OCT-2000 (TREMblrel. 15, Last annotation update)
DE PUTATIVE DNA POLYMERASE III, ALPHA SUBUNIT (EC 2.7.7.7).
GN DNAE OR NMA0632.
OS Neisseria meningitidis (serogroup A).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=65699;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=22491 / SEROGROUP A / SEROTYPE 4A;
RX MEDLINE=20222556; PubMed=10761919;
RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,
RA Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,
RA Davies R.M., Davis P., Devlin K., Feltwell T., Hamlin N., Holroyd S.,
RA Jagers K., Leather S., Moule S., Mungall K., Quail M.A.,
RA Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,
RA Whitehead S., Spratt B.G., Barrall B.G.;
RT "Complete DNA sequence of a serogroup A strain of Neisseria
RT meningitidis 22491."
RL Nature 404:502-506(2000).
DR EMBL: AL162753; CAB83921.1; -.
KW Transferrase; Nucleotidyltransferase.
SQ SEQUENCE 1144 AA; 126975 MW; 95270191E29EBA12 CRC64;

Query Match
Best Local Similarity 83.3%; Score 35; DB 2; Length 1144;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGVLPVG 8
DB 518 GGVLPVG 525

Search completed: June 13, 2001, 14:20:34
Job time: 734 sec

```



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 13, 2001, 14:16:35 ; Search time 118.55 Seconds

(without alignments)  
1.296 Million cell updates/sec

Title: PCT-US01-05825A-15

Perfect score: 42

Sequence: 1 GGVLVQPG 8

Scoring table: BLOSUM62

Searched: Gapop 10.0 , Gapext 0.5

Total number of hits satisfying chosen parameters: 185757

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:\*

1: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep:\*

2: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep:\*

3: /cgn2\_6/ptodata/2/1aa/6A.COMB.pep:\*

4: /cgn2\_6/ptodata/2/1aa/6B.COMB.pep:\*

5: /cgn2\_6/ptodata/2/1aa/PTUS.COMB.pep:\*

6: /cgn2\_6/ptodata/2/1aa/backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	42	100.0	20	2 US-08-859-931A-4	Sequence 4, Appl1
2	39	92.9	118	3 US-08-545-809A-125	Sequence 125, App
3	38	90.5	118	3 US-08-545-809A-108	Sequence 108, App
4	38	90.5	119	2 US-08-475-000-16	Sequence 16, Appl
5	38	90.5	119	2 US-08-483-199-16	Sequence 16, Appl
6	38	90.5	119	2 US-08-484-508-16	Sequence 16, Appl
7	38	90.5	123	2 US-08-428-197-38	Sequence 38, Appl
8	38	90.5	123	5 PCT-US93-10555-38	Sequence 38, Appl
9	35	83.3	15	1 US-08-331-398A-66	Sequence 66, Appl
10	35	83.3	15	2 US-08-652-558-42	Sequence 42, Appl
11	35	83.3	15	2 US-08-331-397B-66	Sequence 66, Appl
12	35	83.3	15	2 US-08-759-804A-65	Sequence 65, Appl
13	35	83.3	18	1 US-08-331-398A-55	Sequence 55, Appl
14	35	83.3	18	2 US-08-401-908-14	Sequence 14, Appl
15	35	83.3	18	2 US-08-331-397B-55	Sequence 55, Appl
16	35	83.3	18	2 US-08-759-804A-54	Sequence 54, Appl
17	35	83.3	20	2 US-08-859-931A-2	Sequence 2, Appl1
18	35	83.3	20	2 US-08-053-451B-113	Sequence 113, App
19	35	83.3	26	1 US-08-471-780C-80	Sequence 80, Appl
20	35	83.3	26	1 US-08-467-282B-80	Sequence 80, Appl
21	35	83.3	26	2 US-08-471-282A-80	Sequence 80, Appl
22	35	83.3	26	2 US-08-466-710C-80	Sequence 80, Appl
23	35	83.3	26	3 US-08-468-739C-80	Sequence 80, Appl
24	35	83.3	29	1 US-08-471-780C-119	Sequence 119, App
25	35	83.3	29	1 US-08-467-282B-119	Sequence 119, App
26	35	83.3	29	2 US-08-471-282A-119	Sequence 119, App
27	35	83.3	29	2 US-08-466-710C-119	Sequence 119, App

28	35	83.3	29	3 US-08-468-739C-119	Sequence 119, App
29	35	83.3	30	1 US-07-988-925-7	Sequence 7, Appl
30	35	83.3	30	1 US-07-977-696C-75	Sequence 75, Appl
31	35	83.3	30	1 US-08-129-930B-75	Sequence 75, Appl
32	35	83.3	30	2 US-08-362-780-7	Sequence 7, Appl
33	35	83.3	30	2 US-08-765-783A-92	Sequence 92, Appl
34	35	83.3	30	2 US-08-470-139-17	Sequence 17, Appl
35	35	83.3	67	1 US-08-162-102C-36	Sequence 36, Appl
36	35	83.3	86	2 US-08-053-451B-126	Sequence 126, App
37	35	83.3	87	1 US-08-497-312-21	Sequence 21, Appl
38	35	83.3	87	2 US-08-765-783A-105	Sequence 105, App
39	35	83.3	98	1 US-08-211-202-116	Sequence 116, App
40	35	83.3	98	1 US-07-942-245-37	Sequence 37, Appl
41	35	83.3	98	2 US-08-428-197-48	Sequence 48, Appl
42	35	83.3	98	2 US-08-665-202-31	Sequence 31, Appl
43	35	83.3	98	5 PCT-US93-10555-48	Sequence 48, Appl
44	35	83.3	108	2 US-08-428-197-4	Sequence 4, Appl1
45	35	83.3	108	5 PCT-US93-10555-4	Sequence 4, Appl1

## ALIGNMENTS

RESULT 1

US-08-859-931A-4

Sequence 4, Application US/08859931A

Patent No. 5945510

GENERAL INFORMATION:

APPLICANT: FASANO, Alessio

TITLE OF INVENTION: SUBSTANTIALLY PURE ZONULIN, A

TITLE OF INVENTION: PHYSIOLOGICAL MODULATOR OF

TITLE OF INVENTION: MAMMALIAN TIGHT JUNCTIONS

NUMBER OF SEQUENCES: 9

CORRESPONDENCE ADDRESS:

ADDRESSEE: SUGHRUE, MIOM, ZINN, MACPEAK & SEAS

STREET: 2100 Pennsylvania Avenue, N.W., Suite 800

City: Washington, D.C.

STATE: D.C.

COUNTRY: U.S.A.

ZIP: 20037

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/859,931A

FILING DATE: 21 MAY 1997

CLASSIFICATION: 514

ATTORNEY/AGENT INFORMATION:

NAME: KIT, Gordon

REGISTRATION NUMBER: 30,764

REFERENCE/DOCKET NUMBER: A-6901

TELECOMMUNICATION INFORMATION:

TELEPHONE: (202) 293-7060

TELEFAX: (202) 293-7860

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 20 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: peptide

HYPOTHETICAL: NO

US-08-859-931A-4

Query Match 100.0%; Score 42; DB 2; Length 20;

Best Local Similarity 100.0%; Pred. NO. 0.095;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGVLVQPG 8

|||||||

Db 8 GGVLVQPG 15

RESULT 2  
US-08-545-809A-125  
Sequence 125, Application US/08545809A  
Patent No. 6096878  
GENERAL INFORMATION:  
APPLICANT: Honjo, Tasuku  
TITLE OF INVENTION: HUMAN IMMUNOGLOBULIN VH GENE  
TITLE OF INVENTION: SEQUENTS AND DNA FRAGMENTS CONTAINING THE SAME  
NUMBER OF SEQUENCES: 145  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson, P.C.  
STREET: 225 Franklin Street  
CITY: Boston  
STATE: MA  
COUNTRY: US  
ZIP: 02110-2804  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: Windows95  
SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/545,809A  
FILING DATE: 27-MAR-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/JP93/00603  
FILING DATE: 10-MAY-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Freeman, John W.  
REGISTRATION NUMBER: 29, 066  
REFERENCE/DOCKET NUMBER: 06501/004001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-542-5370  
TELEFAX: 617-542-8905  
TELEX: 200154  
INFORMATION FOR SEQ ID NO: 125:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 118 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-545-809A-125

Query Match 92.9%; Score 39; DB 3; Length 118;  
Best Local Similarity 87.5%; Pred. No. 2.1;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGAVVQPG 8  
DB 27 GGAVVQPG 34

RESULT 3  
US-08-545-809A-108  
Sequence 108, Application US/08545809A  
Patent No. 6096878  
GENERAL INFORMATION:  
APPLICANT: Honjo, Tasuku  
TITLE OF INVENTION: HUMAN IMMUNOGLOBULIN VH GENE  
TITLE OF INVENTION: SEQUENTS AND DNA FRAGMENTS CONTAINING THE SAME  
NUMBER OF SEQUENCES: 145  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson, P.C.  
STREET: 225 Franklin Street  
CITY: Boston  
STATE: MA  
COUNTRY: US  
ZIP: 02110-2804

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: Windows95  
SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/545,809A  
FILING DATE: 27-MAR-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/JP93/00603  
FILING DATE: 10-MAY-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Freeman, John W.  
REGISTRATION NUMBER: 29, 066  
REFERENCE/DOCKET NUMBER: 06501/004001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-542-5070  
TELEFAX: 617-542-8906  
TELEX: 200154  
INFORMATION FOR SEQ ID NO: 108:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 118 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-545-809A-108

Query Match 90.5%; Score 38; DB 3; Length 118;  
Best Local Similarity 87.5%; Pred. No. 3.1;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGAVVQPG 8  
DB 27 GGAVVQPG 34

RESULT 4  
US-08-475-000-16  
Sequence 16, Application US/08475000  
Patent No. 5811267  
GENERAL INFORMATION:  
APPLICANT: RING, DAVID B.  
TITLE OF INVENTION: ANTIGEN-BINDING SITES OF ANTIBODY  
TITLE OF INVENTION: MOLECULES SPECIFIC FOR CANCER ANTIGENS  
NUMBER OF SEQUENCES: 18  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CHIRON CORPORATION  
STREET: INTELLECTUAL PROPERTY R440 - P.O. BOX 8097  
CITY: Emeryville  
STATE: CA  
COUNTRY: USA  
ZIP: 94662-8097  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC Compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/475,000  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: SAVEREIDE, PAUL  
REGISTRATION NUMBER: 36,914  
REFERENCE/DOCKET NUMBER: 0850.007  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (510) 601-2585  
TELEFAX: (510) 655-3542  
INFORMATION FOR SEQ ID NO: 16:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 119 amino acids  
TYPE: amino acid

TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-475-000-16

Query Match 90.5% Score 38; DB 2; Length 119;  
Best Local Similarity 87.5% Pred. No. 3.2;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GGVLPFG 8  
|||||:  
DB 8 GGVLPFG 15

RESULT 5  
US-08-483-199-16  
Sequence 16, Application US/08483199  
Patent No. 5849877  
GENERAL INFORMATION:

APPLICANT: RING, DAVID B.  
TITLE OF INVENTION: ANTIGEN-BINDING SITES OF ANTIBODY  
TITLE OF INVENTION: MOLECULES SPECIFIC FOR CANCER ANTIGENS  
NUMBER OF SEQUENCES: 18  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CHIRON CORPORATION  
STREET: INTELLECTUAL PROPERTY - R440, P.O. BOX 8097  
CITY: Emeryville  
STATE: CA  
COUNTRY: USA  
ZIP: 94662-8097

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/483.199  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: SAVERIDE, PAUL  
REGISTRATION NUMBER: 36,914  
REFERENCE/DOCKET NUMBER: 0850.009  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (510) 601-2585  
TELEFAX: (510) 655-3542  
INFORMATION FOR SEQ. ID NO.: 16:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 119 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-483-199-16

Query Match 90.5% Score 38; DB 2; Length 119;  
Best Local Similarity 87.5% Pred. No. 3.2;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GGVLPFG 8  
|||||:  
DB 8 GGVLPFG 15

RESULT 6  
US-08-484-508-16  
Sequence 16, Application US/08484508  
Patent No. 5948647  
GENERAL INFORMATION:

APPLICANT: RING, DAVID B.  
TITLE OF INVENTION: ANTIGEN-BINDING SITES OF ANTIBODY  
TITLE OF INVENTION: MOLECULES SPECIFIC FOR CANCER ANTIGENS  
NUMBER OF SEQUENCES: 18

## CORRESPONDENCE ADDRESS:

ADDRESSEE: CHIRON CORPORATION  
STREET: INTELLECTUAL PROPERTY - R440, P.O. BOX 8097  
CITY: Emeryville  
STATE: CA  
COUNTRY: USA  
ZIP: 94662-8097

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/484.508  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:  
NAME: SAVERIDE, PAUL  
REGISTRATION NUMBER: 36,914  
REFERENCE/DOCKET NUMBER: 0850.008  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (510) 601-585  
TELEFAX: (510) 655-3542  
INFORMATION FOR SEQ. ID NO.: 16:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 119 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-484-508-16

US-08-484-508-16

Query Match 90.5% Score 38; DB 2; Length 119;  
Best Local Similarity 87.5% Pred. No. 3.2;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GGVLPFG 8  
|||||:  
DB 8 GGVLPFG 15

RESULT 7  
US-08-428-197-38  
Sequence 38, Application US/08428197  
Patent No. 3891438  
GENERAL INFORMATION:

APPLICANT: SILVERMAN, GREGG J.  
TITLE OF INVENTION: METHOD FOR STIMULATING PRODUCTION OF  
TITLE OF INVENTION: VARIABLE REGION GENE FAMILY RESTRICTED ANTIBODIES THROUGH  
TITLE OF INVENTION: VACCINATION WITH A B-CELL SUPERNATANT AND CONJUGATES  
NUMBER OF SEQUENCES: 51  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Spensley Horn Jupas & Lubitz  
STREET: 1880 Century Park East - Suite 500  
CITY: Los Angeles  
STATE: California  
COUNTRY: USA  
ZIP: 90067

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/428.197  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US93/10555  
FILING DATE: 29-OCT-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Howells, Stacy L.

NAME: Howells, Stacy L.

REGISTRATION NUMBER: 34,842  
REFERENCE/DOCKET NUMBER: FD-2630  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 455-5100  
TELEFAX: (619) 455-5110  
INFORMATION FOR SEQ ID NO: 38:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 123 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-428-197-38

Query Match 90.5%; Score 38; DB 2; Length 123;  
Best Local Similarity 87.5%; Pred. No. 3.3;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
OY 1 GGVLPQPG 8  
Db 8 GGVLPQPG 15

RESULT 8  
PCT-US93-10555-38  
Sequence 38, Application 10/US9310555  
GENERAL INFORMATION:  
APPLICANT: SILVERMAN, GREGG J.  
TITLE OF INVENTION: METHOD FOR STIMULATING PRODUCTION OF  
TITLE OF INVENTION: VARIABLE REGION GENE FAMILY RESTRICTED ANTIBODIES THROUGH  
TITLE OF INVENTION: VACCINATION WITH A B-CELL SUPERANTIGEN AND CONJUGATES  
NUMBER OF SEQUENCES: 51  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Spensley Horn Jubas & Lubitz  
STREET: 1880 Century Park East - Suite 500  
CITY: Los Angeles  
STATE: California  
COUNTRY: USA  
ZIP: 90067  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US93/10555  
FILING DATE: 29-OCT-1993  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Howells, Stacy L.  
REGISTRATION NUMBER: 34,842  
REFERENCE/DOCKET NUMBER: FD-2630  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 455-5100  
TELEFAX: (619) 455-5110  
INFORMATION FOR SEQ ID NO: 38:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 123 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
PCT-US93-10555-38

Query Match 90.5%; Score 38; DB 5; Length 123;  
Best Local Similarity 87.5%; Pred. No. 3.3;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
OY 1 GGVLPQPG 8  
Db 8 GGVLPQPG 15

RESULT 9  
US-08-331-398A-66  
Sequence 66, Application US/08331398A  
Patent No. 5608039  
GENERAL INFORMATION:  
APPLICANT: Pastan, Ira  
APPLICANT: Willingham, Mark  
APPLICANT: Fitzgerald, David  
APPLICANT: Brinkmann, Ulrich  
APPLICANT: Pai, Lee  
TITLE OF INVENTION: Single Chain B3 Antibody Fusion Proteins  
TITLE OF INVENTION: and Their Uses (as amended)  
NUMBER OF SEQUENCES: 68  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend and Crew  
STREET: One Market Plaza, Stewart Street Plaza  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94105-1492  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/331,398A  
FILING DATE: 28-OCT-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
PRIOR APPLICATION NUMBER: US 07/767,331  
FILING DATE: 30-SEP-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/596,289  
FILING DATE: 12-OCT-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: Hunter, Tom  
REGISTRATION NUMBER: 38,498  
REFERENCE/DOCKET NUMBER: 015280-126110US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 543-9600  
TELEFAX: (415) 543-5043  
INFORMATION FOR SEQ ID NO: 66:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 15 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-331-398A-66

Query Match 83.3%; Score 35; DB 1; Length 15;  
Best Local Similarity 87.5%; Pred. No. 1.3;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
OY 1 GGVLPQPG 8  
Db 8 GGVLPQPG 15

RESULT 10  
US-08-652-558-42  
Sequence 42, Application US/08652558  
Patent No. 5861155  
GENERAL INFORMATION:  
APPLICANT: LIN, AUGUSTINE YEE-THARN  
TITLE OF INVENTION: HUMANIZED ANTIBODIES AND USES  
TITLE OF INVENTION: THEREOF  
NUMBER OF SEQUENCES: 52  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BANNER & WITCOFF

STREET: 75 STATE STREET, 23RD FLOOR  
CITY: BOSTON  
STATE: MASSACHUSETTS  
COUNTRY: USA  
ZIP: 02109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 inch diskette, 1.4 Mb storage  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: WordPerfect 6.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/652,558  
FILING DATE: JUNE 6, 1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/IB94/00387  
FILING DATE: NOVEMBER 21, 1994  
ATTORNEY/AGENT INFORMATION:  
NAME: YANKWICH, LEON R.  
REGISTRATION NUMBER: 30,237  
REFERENCE/DOCKET NUMBER: 95,497-L  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-345-9111  
TELEFAX: 617-345-9100  
INFORMATION FOR SEQ ID NO: 42:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 15 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-652-558-42

Query Match 83.3%; Score 35; DB 2; Length 15;  
Best Local Similarity 87.5%; Pred. No. 1.3;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 GGVLPQPG 8  
|||  
Db 8 GGVLPQPG 15

RESULT 11  
US-08-331-397B-66  
Sequence 66 Application US/08331397B  
Patent No. 5981726  
GENERAL INFORMATION:  
APPLICANT: Pastan, Ira  
TITLE OF INVENTION: Chimeric and Mutationally Stabilized Tumor-  
TITLE OF INVENTION: Specific Antibody Fragments, Fusion Proteins, and Uses  
TITLE OF INVENTION: Thereof  
NUMBER OF SEQUENCES: 68  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Townsend and Townsend and Crew  
STREET: One Market Plaza, Stewart Street Plaza  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94105-1492  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/331,397B  
FILING DATE: 28-OCT-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/767,331  
FILING DATE: 30-SEP-1991  
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/596,289  
FILING DATE: 12-OCT-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: Hunter, Tom  
REGISTRATION NUMBER: 38,498  
REFERENCE/DOCKET NUMBER: 015280-126120US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 543-9600  
TELEFAX: (415) 543-5043  
INFORMATION FOR SEQ ID NO: 66:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 15 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-331-397B-66

Query Match 83.3%; Score 35; DB 2; Length 15;  
Best Local Similarity 87.5%; Pred. No. 1.3;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 GGVLPQPG 8  
|||  
Db 8 GGVLPQPG 15

RESULT 12  
US-08-759-804A-65  
Sequence 65 Application US/08759804A  
Patent No. 5990296  
GENERAL INFORMATION:  
APPLICANT: Pastan, Ira  
APPLICANT: Willingham, Mark  
APPLICANT: Fitzgerald, David J.  
APPLICANT: Brinkmann, Ulrich  
APPLICANT: Pai, Lee  
TITLE OF INVENTION: Tumor-Specific Antibody Fragments,  
TITLE OF INVENTION: Fusion Proteins, and Uses Thereof  
NUMBER OF SEQUENCES: 68  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Center, Eighth Floor  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-3834  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/759,804A  
FILING DATE: 03-DEC-1996  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/331,398  
FILING DATE: 28-OCT-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/767,331  
FILING DATE: 30-SEP-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/596,289  
FILING DATE: 12-OCT-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: Weber, Ellen L.  
REGISTRATION NUMBER: 32,762  
REFERENCE/DOCKET NUMBER: 015280-126140US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 576-0200  
TELEFAX: (415) 576-0300

INFORMATION FOR SEQ ID NO: 65:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 15 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-759-804A-65

Query Match 83.3%; Score 35; DB 2; Length 15;  
Best Local Similarity 87.5%; Pred. No. 1.3;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGVLVOPG 8  
111111  
Db 8 GGGLVOPG 15

RESULT 13  
US-08-331-398A-55  
Sequence 55, Application US/08331398A  
Patent No. 5608039  
GENERAL INFORMATION:  
APPLICANT: Pastan, Ira  
APPLICANT: Willingham, Mark  
APPLICANT: Fitzgerald, David  
APPLICANT: Brinkmann, Ulrich  
APPLICANT: Pal, Lee  
TITLE OF INVENTION: Single Chain B3 Antibody Fusion Proteins  
TITLE OF INVENTION: and Their Uses (as amended)  
NUMBER OF SEQUENCES: 68  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend and Crew  
STREET: One Market Plaza, Stewart Street Plaza  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94105-1492  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/331,398A  
FILING DATE: 28-OCT-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/67,731  
FILING DATE: 30-SEP-1991  
APPLICATION NUMBER: US 07/596,289  
FILING DATE: 12-OCT-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: Hunter, Tom  
REGISTRATION NUMBER: 38,498  
REFERENCE/DOCKET NUMBER: 015280-126110US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 543-9600  
TELEFAX: (415) 543-5043  
INFORMATION FOR SEQ ID NO: 55:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 18 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-331-398A-55

Query Match 83.3%; Score 35; DB 1; Length 18;  
Best Local Similarity 87.5%; Pred. No. 1.6;

Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 GGVLVOPG 8  
111111  
Db 8 GGGLVOPG 15

RESULT 14  
US-08-401-908-14  
Sequence 14, Application US/08401908  
Patent No. 568146  
GENERAL INFORMATION:  
APPLICANT: Yoshitoku OKUNO et al.  
TITLE OF INVENTION: A DNA WHICH CODES FOR THE VARIABLE REGION OF  
NUMBER OF SEQUENCES: 36  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Wenderoth, Lind & Ponack  
STREET: 805 Fifteenth Street, N.W., #700  
CITY: Washington  
STATE: D.C.  
COUNTRY: U.S.A.  
ZIP: 20005  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: Wordperfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/401,908  
FILING DATE: March 10, 1995  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Warren M. Cheek, Jr.  
REGISTRATION NUMBER: 33,367  
REFERENCE/DOCKET NUMBER:  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-371-8850  
TELEFAX:  
TELEX:  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 18  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
ORIGINAL SOURCE: heavy chain of C179  
US-08-401-908-14

Query Match 83.3%; Score 35; DB 1; Length 18;  
Best Local Similarity 87.5%; Pred. No. 1.6;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGVLVOPG 8  
111111  
Db 8 GGGLVOPG 15

RESULT 15  
US-08-331-397B-55  
Sequence 55, Application US/08331397B  
Patent No. 5981726  
GENERAL INFORMATION:  
APPLICANT: Pastan, Ira  
APPLICANT: Benhar, Itai  
TITLE OF INVENTION: Chimeric and Mutationally Stabilized Tumor-Specific Antibody Fragments, Fusion Proteins, and Uses Thereof



NUMBER OF SEQUENCES: 68  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend and Crew  
STREET: One Market Plaza, Steuart Street Plaza  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94105-1492  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/331,397B  
FILING DATE: 28-OCT-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/767,331  
FILING DATE: 30-SEP-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/596,289  
FILING DATE: 12-OCT-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: Hunter, Tom  
REGISTRATION NUMBER: 38,498  
REFERENCE/DOCKET NUMBER: 015280-126120US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 543-9600  
TELEFAX: (415) 543-5043  
INFORMATION FOR SEQ ID NO: 55:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 18 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-331-397B-55

Query Match 83.3%; Score 35; DB 2; Length 18;  
Best Local Similarity 87.5%; Pred. No. 1.6;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGVLVQPG 8  
|||  
Db 8 GGVLVQPG 15

Search completed: June 13, 2001, 14:16:36  
Job time: 497 sec

*This Page Blank (uspto)*

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OW protein - protein search, using sw model

Run on: June 13, 2001, 14:14:38 ; Search time 229.28 Seconds  
(without alignments)

1.995 Million cell updates/sec

Title: PCT-US01-05825A-16

Perfect score: 41

Sequence: 1 GGVLVODG 8

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 390729 seqs, 57163235 residues

Total number of hits satisfying chosen parameters: 390729

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A.Geneseq\_0401.\*  
1: /SID6/gcgdata/geneseq/geneseq/AA1980.DAT.\*  
2: /SID6/gcgdata/geneseq/geneseq/AA1981.DAT.\*  
3: /SID6/gcgdata/geneseq/geneseq/AA1982.DAT.\*  
4: /SID6/gcgdata/geneseq/geneseq/AA1983.DAT.\*  
5: /SID6/gcgdata/geneseq/geneseq/AA1984.DAT.\*  
6: /SID6/gcgdata/geneseq/geneseq/AA1985.DAT.\*  
7: /SID6/gcgdata/geneseq/geneseq/AA1986.DAT.\*  
8: /SID6/gcgdata/geneseq/geneseq/AA1987.DAT.\*  
9: /SID6/gcgdata/geneseq/geneseq/AA1988.DAT.\*  
10: /SID6/gcgdata/geneseq/geneseq/AA1989.DAT.\*  
11: /SID6/gcgdata/geneseq/geneseq/AA1990.DAT.\*  
12: /SID6/gcgdata/geneseq/geneseq/AA1991.DAT.\*  
13: /SID6/gcgdata/geneseq/geneseq/AA1992.DAT.\*  
14: /SID6/gcgdata/geneseq/geneseq/AA1993.DAT.\*  
15: /SID6/gcgdata/geneseq/geneseq/AA1994.DAT.\*  
16: /SID6/gcgdata/geneseq/geneseq/AA1995.DAT.\*  
17: /SID6/gcgdata/geneseq/geneseq/AA1996.DAT.\*  
18: /SID6/gcgdata/geneseq/geneseq/AA1997.DAT.\*  
19: /SID6/gcgdata/geneseq/geneseq/AA1998.DAT.\*  
20: /SID6/gcgdata/geneseq/geneseq/AA1999.DAT.\*  
21: /SID6/gcgdata/geneseq/geneseq/AA2000.DAT.\*  
22: /SID6/gcgdata/geneseq/geneseq/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result NO.	Score	Query Match	Length	DB ID	Description
1	41	100.0	8	21	Y79120
2	38	92.7	8	21	Y79124
3	36	87.8	8	21	Y79118
4	36	87.8	115	20	Y42459
5	36	87.8	115	20	Y42460
6	35	85.4	115	20	Y27892
7	34	82.9	8	21	Y84661
8	34	82.9	8	21	Y79119
9	34	82.9	8	21	Y79128
10	34	82.9	20	20	W94489
11	34	82.9	20	21	Y79133

12	34	82.9	25	15	R52775	Murine BRE-3 immun
13	34	82.9	25	15	R52793	Murine BRE-3 immun
14	33	80.5	8	21	Y79108	Peptide antagonist
15	33	80.5	8	21	Y79122	Peptide antagonist
16	33	80.5	166	21	G49731	Arabidopsis thaliana
17	33	80.5	355	20	Y32173	Chlamydia psittaci
18	33	80.5	368	16	R67970	Alfalfa beta-1,3-g
19	32	78.0	92	21	G49732	Arabidopsis thaliana
20	32	78.0	401	19	W55033	Actinobacillus ple
21	32	78.0	401	20	Y21970	Banana ripening fr
22	32	78.0	1187	20	Y05841	Anti-HIL2 antilbod
23	31	75.6	8	21	Y79123	Human OREF ORF899
24	31	75.6	98	21	B40074	Human OREF ORF899
25	31	75.6	108	21	B41135	Human immunoglobul
26	31	75.6	112	16	R66331	Vh region from an
27	31	75.6	122	16	R80271	Amino acid sequenc
28	31	75.6	124	19	W80816	Arabidopsis thaliana
29	31	75.6	140	21	G06871	Human gastric can
30	31	75.6	176	22	B63642	Arabidopsis thaliana
31	31	75.6	196	20	Y36831	Human acid sequenc
32	31	75.6	197	21	G06870	Arabidopsis thaliana
33	31	75.6	212	16	R88280	Arabidopsis thaliana
34	31	75.6	212	21	Y51026	A. gossypii rib3 p
35	31	75.6	212	21	Y51029	A. gossypii rib3 p
36	31	75.6	384	19	W75258	Fragment of human
37	31	75.6	529	19	W82586	Human Patched-2 pr
38	31	75.6	856	19	W52221	Human secreted pro
39	31	75.6	1146	21	Y92225	Human patched-2 (p
40	31	75.6	1203	20	Y43261	Human patched-2 pr
41	31	75.6	1203	20	Y28444	Human patched-2 pr
42	31	75.6	1203	21	Y92703	Human patched-2 (p
43	30	73.2	8	21	Y79112	Peptide antagonist
44	30	73.2	102	20	W93352	Human adult testis
45	30	73.2	115	20	Y42461	Rat guanlylin prepr

#### ALIGNMENTS

RESULT 1	
ID	Y79120 standard; Peptide; 8 AA.
XX	
AC	Y79120;
XX	
DT	05-JUN-2000 (first entry)
XX	
DE	Peptide antagonist of zonulin.
XX	
KW	zonulin; antagonist; zonula occludens toxin receptor;
KW	blood-brain barrier; antiinflammatory; cerebroprotective;
KW	neuroprotective; dermatological; antitumor; antiviral;
KW	antibacterial; cytostatic; anti-HIV; vulnery; antiallergic;
KW	hypotensive; immunosuppressive; antiparasitic; vasotropic;
KW	gastrointestinal inflammation; therapy.
XX	
OS	Synthetic.
XX	
PN	WO200007609-A1.
XX	
PD	17-FEB-2000.
XX	
PF	28-JUL-1999; 99WO-US16683.
XX	
PR	03-AUG-1998; 98US-0127815.
XX	
PA	(UYMA-) UNIV MARYLAND BALTIMORE.
XX	
PI	Fasano A;
XX	
DR	WPI, 2000-205565/18.
XX	
PT	New peptide antagonist of zonulin useful as antiinflammatory agent for

PI treating cerebral ischemia, stroke, cerebral edema, gastritis,  
 PT shigellosis, viral gastroenteritis, meningitis, encephalomyelitis  
 XX  
 PS Claim 1; Page 46; 69pp; English.

CC This present sequence is that of a peptide antagonist of zonulin  
 CC (2), one of 25 such peptides (see Y79105-29) of the invention,  
 CC which bind to a zonula occludens toxin (ZOT) receptor, yet do not  
 CC physiologically modulate the opening of mammalian tight junctions  
 CC (7). The peptide antagonists are based on a common motif of ZOT  
 CC and human zonulins, which is believed to be critical for receptor  
 CC binding. They can be prepared by chemical syntheses or by use of  
 CC recombinant DNA techniques. The peptide antagonists are used as an  
 CC antiinflammatory agents in the treatment of gastrointestinal  
 CC inflammation, where they bind to the ZOT receptor in the intestine  
 CC and yet does not physiologically modulate the opening of TJ in the  
 CC intestine. Gastrointestinal inflammation conditions give rise to  
 CC increased intestinal permeability and the peptide is useful for  
 CC treating intestinal conditions that cause protein losing enteropathy  
 CC caused by infection, e.g. Clostridium difficile infection,  
 CC enterocolitis, shigellosis, viral gastroenteritis, parasite  
 CC infestation, bacterial overgrowth, whipple's disease, diseases with  
 CC mucosal erosion or ulcerations, e.g. gastritis, gastric cancer,  
 CC collagenous colitis, inflammatory bowel disease, diseases marked by  
 CC lymphatic obstruction, e.g. congenital intestinal lymphangiectasia,  
 CC sarcoidosis lymphoma, mesenteric tuberculosis, and after surgical  
 CC correction of congenital heart disease with Fontan's operation,  
 CC mucosal diseases without ulceration, e.g. Menetrier's disease,  
 CC coeliac disease, eosinophilic gastroenteritis, and immune diseases,  
 CC e.g. systemic lupus erythematosus or food allergies, primarily to  
 CC milk.

XX Sequence 8 AA:

Query Match 100.0%; Score 41; DB 21; Length 8;  
 Best Local Similarity 100.0%; Pred. No. 3.2e+05;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GGVLVODG 8  
 ||:|||||  
 Db 1 ggvlvodg 8

# RESULT 2

ID Y79124 standard; Peptide; 8 AA.

XX Y79124;

DT 05-JUN-2000 (first entry)

DE Peptide antagonist of zonulin.

XX Zonulin; antagonist; zonula occludens toxin receptor;  
 KW blood-brain barrier; antiinflammatory; cerebroprotective;  
 KW neuroprotective; dermatological; anticancer; antiviral;  
 KW antibacterial; cytostatic; anti-HIV; vulnerary; antiallergic;  
 KW hypotensive; immunosuppressive; antiparasitic; vasotropic;  
 KW gastrointestinal inflammation; therapy.

OS Synthetic.

PN WO200007609-A1.

PD 17-FEB-2000.

PF 28-JUL-1999; 99WO-US16683.

PR 03-AUG-1998; 98US-0127815.

PA (UYMA-) UNIV MARYLAND BALTIMORE.

PI Fasano A;  
 XX WPI; 2000-205565/18.

PT New peptide antagonist of zonulin useful as antiinflammatory agent for  
 PT treating cerebral ischemia, stroke, cerebral edema, gastritis,  
 PT shigellosis, viral gastroenteritis, meningitis, encephalomyelitis  
 XX

PS Claim 1; Page 47; 69pp; English.

CC This present sequence is that of a peptide antagonist of zonulin  
 CC (2), one of 25 such peptides (see Y79105-29) of the invention,  
 CC which bind to a zonula occludens toxin (ZOT) receptor, yet do not  
 CC physiologically modulate the opening of mammalian tight junctions  
 CC (7). The peptide antagonists are based on a common motif of ZOT  
 CC and human zonulins, which is believed to be critical for receptor  
 CC binding. They can be prepared by chemical syntheses or by use of  
 CC recombinant DNA techniques. The peptide antagonists are used as an  
 CC antiinflammatory agents in the treatment of gastrointestinal  
 CC inflammation, where they bind to the ZOT receptor in the intestine  
 CC and yet does not physiologically modulate the opening of TJ in the  
 CC intestine. Gastrointestinal inflammation conditions give rise to  
 CC increased intestinal permeability and the peptide is useful for  
 CC treating intestinal conditions that cause protein losing enteropathy  
 CC caused by infection, e.g. Clostridium difficile infection,  
 CC enterocolitis, shigellosis, viral gastroenteritis, parasite  
 CC infestation, bacterial overgrowth, whipple's disease, diseases with  
 CC mucosal erosion or ulcerations, e.g. gastritis, gastric cancer,  
 CC collagenous colitis, inflammatory bowel disease, diseases marked by  
 CC lymphatic obstruction, e.g. congenital intestinal lymphangiectasia,  
 CC sarcoidosis lymphoma, mesenteric tuberculosis, and after surgical  
 CC correction of congenital heart disease with Fontan's operation,  
 CC mucosal diseases without ulceration, e.g. Menetrier's disease,  
 CC coeliac disease, eosinophilic gastroenteritis, and immune diseases,  
 CC e.g. systemic lupus erythematosus or food allergies, primarily to  
 CC milk.

XX Sequence 8 AA:

Query Match 92.7%; Score 38; DB 21; Length 8;  
 Best Local Similarity 87.5%; Pred. No. 3.2e+05;  
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GGVLVODG 8  
 ||:|||||  
 Db 1 ggvlvodg 8

# RESULT 3

ID Y79118 standard; Peptide; 8 AA.

XX Y79118;

DT 05-JUN-2000 (first entry)

DE Peptide antagonist of zonulin.

XX Zonulin; antagonist; zonula occludens toxin receptor;  
 KW blood-brain barrier; antiinflammatory; cerebroprotective;  
 KW neuroprotective; dermatological; anticancer; antiviral;  
 KW antibacterial; cytostatic; anti-HIV; vulnerary; antiallergic;  
 KW hypotensive; immunosuppressive; antiparasitic; vasotropic;  
 KW gastrointestinal inflammation; therapy.

OS Synthetic.

PN WO200007609-A1.

PD 17-FEB-2000.

PF 28-JUL-1999; 99WO-US16683.

[illegible]

XX	PN	US5969097-A.
XX	PD	19-OCT-1999.
XX	PE	23-JUN-1992; 92US-0903029.
XX	PR	23-JUN-1992; 92US-0903029.
XX	PA	(SEAR ) SEARLE & CO G D.
XX	PI	Wiegand RC, Currie MG, Fok KF;
XX	DR	WPI: 1999-590442/50.
XX	N-PSDB:	Z22677.
PT	Isolated protein used as a laxative in the treatment of constipation -	
XX	Disclosure: Fig 1; 14pp; English.	
XX	This sequence represents a human guanylin preprohormone which is	
CC	cleaved to generate a 15 amino acid mature peptide (Y42458). This	
CC	sequence is the encoded protein from a composite DNA sequence isolated	
CC	by probing a human duodenum cDNA library with the corresponding rat	
CC	coding sequence. Guanylin is an intestinal guanylate cyclase regulator	
CC	which has similarity to bacterial heat stable enterotoxins which	
CC	activate intestinal guanylate cyclases. The guanylin peptide can be	
XX	used as a laxative in the treatment of constipation.	
SQ	Sequence 115 AA;	
OY	1 GGVTVDG 8         Db 20 ggvvtvdg 27.	
RESULT 5		
Y42460		
ID Y42460	standard; Protein; 115 AA.	
AC Y42460;		
DT 21-DEC-1999	(first entry)	
DE Human guanylin preprohormone sequence.		
XX Human; guanylin; preprohormone; intestinal guanylate cyclase regulator;		
KW bacterium; heat stable enterotoxin; laxative; constipation.		
RW Homo sapiens.		
XX US5969097-A.		
PN 19-OCT-1999.		
PD 23-JUN-1992; 92US-0903029.		
PF 23-JUN-1992; 92US-0903029.		
PR 23-JUN-1992; 92US-0903029.		
XX (SEAR ) SEARLE & CO G D.		
PA Wiegand RC, Currie MG, Fok KF;		
PI WPI: 1999-590442/50.		
DR Isolated protein used as a laxative in the treatment of constipation -		
XX Disclosure: Fig 2; 14pp; English.		

XX This sequence represents a human guanylin preprohormone which is cleaved  
CC to generate a 15 amino acid mature peptide (Y42458). Guanylin is an  
CC intestinal guanylate cyclase regulator which has similarity to bacterial  
CC heat stable enterotoxins which activate intestinal guanylate cyclases.  
CC The guanylin peptide can be used as a laxative in the treatment of  
CC constipation.  
XX  
SQ Sequence 115 AA;  
Query Match 87.8%; Score 36; DB 20; Length 115;  
Best Local Similarity 87.5%; Pred. No. 12;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
OY 1 GGVLVQDG 8  
||| |||  
Db 20 ggvtvqdg 27  
RESULT 6  
Y27892 6  
ID Y27892 standard; Protein: 115 AA.  
AC Y27892;  
XX 30-JUL-1999 (first entry)  
DE Human secreted protein encoded by gene No. 116.  
XX  
KW Human; secreted protein; fusion protein; gene therapy; protein therapy;  
KW diagnosis; tissue; cancer; neurodegenerative disorder; leukaemia;  
KW developmental abnormality; foetal deficiency; blood; allergy; renal;  
KW immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;  
KW inflammation; ischemic shock; Alzheimer's disease; restenosis; AIDS;  
KW cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;  
KW osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;  
KW endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.  
XX  
OS Homo sapiens.  
XX  
PN M09924836-A1.  
XX 20-MAY-1999.  
XX  
XX 04-NOV-1998; 98MO-US23435.  
XX  
PR 17-NOV-1997; 97US-0056100.  
PR 07-NOV-1997; 97US-0054900.  
PR 07-NOV-1997; 97US-0054908.  
PR 07-NOV-1997; 97US-0054911.  
PR 07-NOV-1997; 97US-0054912.  
PR 07-NOV-1997; 97US-0054983.  
PR 07-NOV-1997; 97US-0054984.  
PR 07-NOV-1997; 97US-0064885.  
PR 07-NOV-1997; 97US-0064887.  
PR 07-NOV-1997; 97US-0064988.  
PR 17-NOV-1997; 97US-0066090.  
PR 17-NOV-1997; 97US-0066094.  
PR 17-NOV-1997; 97US-0066095.  
PR 17-NOV-1997; 97US-0066089.  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
PI Carter KC, Ebner R, Endress GA, Feng P, Janat F;  
PI Kyaw H, Lafleur DW, Moore PA, Ni J, Olsen HS, Rosen CA;  
PI Ruben SM, Shi Y, Soppet DR, Wei Y;  
XX  
DR WPI: 1999-337740/28.  
DR N-PSDB: X85048.  
XX  
PT New human secreted proteins and coding sequences useful for treating  
PT disorders of the immune system and hyperproliferative disorders

XX Disclosure: Page 166; 507pp; English.  
PS  
XX This sequence represents a secreted human protein encoded by the gene  
XX clone detailed in the descriptor line. The gene can be used to generate  
XX fusion proteins by linking to the gene to a human immunoglobulin Fc  
XX portion (e.g. X84924) for increasing the stability of the fused protein  
XX as compared to the human protein only.  
CC The invention relates to 125 novel genes and their fragments (nucleic  
CC acid sequences: X84933-X85057; amino acid sequences Y27567-Y27933) which  
CC are useful for preventing, treating or ameliorating medical conditions  
CC e.g. by protein or gene therapy. Also, pathological conditions can be  
CC diagnosed by determining the amount of the new polypeptides in a sample  
CC or by determining the presence of mutations in the new polynucleotides.  
CC Specific uses are described for each of the 125 polynucleotides, based on  
CC which tissues they are most highly expressed in (see X84933 for described  
CC uses).  
XX  
SQ Sequence 115 AA;  
Query Match 85.4%; Score 35; DB 20; Length 115;  
Best Local Similarity 87.5%; Pred. No. 19;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
OY 1 GGVLVQDG 8  
||| |||  
Db 1 ggvtvqdg 8  
RESULT 7  
Y84661 7  
ID Y84661 standard; Protein: 8 AA.  
AC Y84661;  
XX 25-JUL-2000 (first entry)  
DE Peptide antagonist Fz1/0 of zonula occludens toxin (zot) polypeptide.  
XX  
XX Human; zot; zonula occludens toxin; zonulin; antigen presenting cell;  
KW APC; lymphocyte proliferation; antigen; auto-immune disorder;  
KW immune-related disorder; immune system rejection; multiple sclerosis;  
KW organ transplantation; inflammatory disease; allergic disease;  
KW rheumatoid arthritis; insulin dependent diabetes mellitus;  
KW celiac disease; Sjogren's syndrome; systemic lupus erythematosus;  
KW auto-immune thyroiditis; idiopathic thrombocytopenic purpura;  
KW hemolytic anemia; Grave's disease; Addison disease; autoimmune orchitis;  
KW pernicious anemia; vasculitis; autoimmune coagulopathy; polymyositis;  
KW myasthenia gravis; polynephritis; pemphigus; rheumatic carditis;  
KW dermatomyositis; scleroderma; asthma; psoriasis; eczematous dermatitis;  
KW Kaposi's sarcoma; inflammatory bowel disease; proliferative disorder;  
KW PCR primer; ss.  
XX  
OS Synthetic.  
XX  
PN W0200015252-A1.  
XX 23-MAR-2000.  
XX  
XX 09-SEP-1999; 99MO-US18842.  
XX  
XX 14-SEP-1998; 98US-0100266.  
XX  
PA (UYMA-) UNIV MARYLAND BALTIMORE.  
XX  
PI Fasano A, Seteln MB, Lu R, Tanner MK;  
PI  
XX  
DR WPI: 2000-271257/23.  
XX  
PT Suppression of antigen presenting cell mediated lymphocyte  
PT proliferation, by administering a zot-related immunoregulator useful  
PT for treating immune-related disorders, immune system rejection

PT subsequent to tissue or organ transplantation  
XX  
PS Example 7; Page 59; 95pp; English.  
XX  
CC The present sequence represents a peptide antagonist of zonula occludens  
CC toxin (zot) polypeptide. The specification describes a method of  
CC suppressing antigen presenting cell (APC)-mediated lymphocyte  
CC proliferation in a mammalian host pre-exposed to a particular antigen.  
CC The method comprises administering to the host an effective amount of a  
CC zot-related immunoregulator selected from zot (zonula occludens toxin) or  
CC zonulin, the amount effective to down-regulate the activity of the APC.  
CC The method can be used to down-regulate APC-mediated lymphocyte  
CC proliferation in mammalian hosts suffering from auto-immune or  
CC immune-related disorders, immune system rejection subsequent to tissue  
CC or organ transplantation, or inflammatory or allergic diseases. The  
CC autoimmune or immune related disorders include multiple sclerosis,  
CC rheumatoid arthritis, insulin dependent diabetes mellitus, celiac  
CC disease, Sjogren's syndrome, systemic lupus erythematosus, auto-immune  
CC thyroiditis, idiopathic thrombocytopenic purpura, hemolytic anemia,  
CC Grave's disease, Addison disease, autoimmune orchitis, pernicious  
CC anemia, vasculitis, autoimmune coagulopathies, myasthenia gravis,  
CC polynueritis, pemphigus, rheumatic carditis, polymyositis,  
CC dermatomyositis, and scleroderma. The inflammatory or allergic disease  
CC or disorder is selected from asthma, psoriasis, eczematous dermatitis,  
CC Kaposi's sarcoma, multiple sclerosis, inflammatory bowel disease,  
CC proliferative disorders of smooth muscle cells, and inflammatory  
CC conditions associated with mycotic, viral, parasitic, or bacterial  
CC infections.  
CC  
SQ Sequence 8 AA;  
  
Query Match 82.9%; Score 34; DB 21; Length 8;  
Best Local Similarity 87.5%; Pred. No. 3.2e+05;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
  
QY 1 GGVIVQDG 8  
Db 1 ggvivqpg 8  
|||||  
|  
  
RESULT 8  
Y79119  
ID Y79119 standard; Peptide: 8 AA.  
XX  
AC Y79119;  
XX  
DT 05-JUN-2000 (first entry)  
XX  
DE Peptide antagonist of zonulin.  
XX  
KW Zonulin; antagonist; zonula occludens toxin receptor; human;  
KW blood-brain barrier; antiinflammatory; cerebroprotective;  
KW neuroprotective; dermatological; antitumor; antiviral;  
KW antibacterial; cytostatic; anti-HIV; vulnerrary; antiallergic;  
KW hypotensive; immunosuppressive; antiparasitic; vasotropic;  
KW gastrointestinal inflammation; therapy.  
XX  
OS Homo sapiens.  
XX  
PN WO200007609-A1.  
XX  
PD 17-FEB-2000.  
XX  
PE 28-JUL-1999; 99WO-US16683.  
XX  
PR 03-AUG-1998; 98US-0127815.  
XX  
PA (UYMA-) UNIV MARYLAND BALTIMORE.  
XX  
PI Faasano A;  
XX  
DR WPI; 2000-205565/18.

XX  
PT New peptide antagonist of zonulin useful as antiinflammatory agent for  
PT treating cerebral ischemia, stroke, cerebral edema, gastritis,  
PT shigellosis, viral gastroenteritis, meningitis, encephalomyelitis  
XX  
PS Claim 1; Page 45; 69pp; English.  
XX  
CC This present sequence is that of a peptide antagonist of zonulin,  
CC corresponding to residues 8-15 of human foetal intestinal zonulin.  
CC It is one of 25 peptide antagonists (see Y79105-29) of the invention  
CC that bind to a zonula occludens toxin (ZOT) receptor, yet do not  
CC physiologically modulate the opening of mammalian tight junctions  
CC (TJ). The peptide antagonists are based on a common motif of ZOT  
CC and human zonulins, which is believed to be critical for receptor  
CC binding. They can be prepared by chemical syntheses or by use of  
CC recombinant DNA techniques. The peptide antagonists are used as an  
CC antiinflammatory agents in the treatment of gastrointestinal  
CC inflammation, where they bind to the ZOT receptor in the intestine  
CC and yet does not physiologically modulate the opening of TJ in the  
CC intestine. Gastrointestinal inflammation conditions give rise to  
CC increased intestinal permeability and the peptide is useful for  
CC treating intestinal conditions that cause protein losing enteropathy  
CC caused by infection, e.g. Clostridium difficile infection,  
CC enterocolitis, shigellosis, viral gastroenteritis, parasite  
CC infestation, bacterial overgrowth, whipple's disease, diseases with  
CC mucosal erosion or ulcerations, e.g. gastritis, gastric cancer,  
CC collagenous colitis, inflammatory bowel disease, diseases marked by  
CC lymphatic obstruction, e.g. congenital intestinal lymphangiectasia,  
CC sarcoidosis lymphoma, mesenteric tuberculosis, and after surgical  
CC correction of congenital heart disease with Fontan's operation,  
CC mucosal diseases without ulceration, e.g. Menetrier's disease,  
CC coeliac disease, eosinophilic gastroenteritis, and immune diseases,  
CC e.g. systemic lupus erythematosus or food allergies, primarily to  
CC milk.  
CC  
SQ Sequence 8 AA;  
  
Query Match 82.9%; Score 34; DB 21; Length 8;  
Best Local Similarity 87.5%; Pred. No. 3.2e+05;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
  
QY 1 GGVIVQDG 8  
Db 1 ggvivqpg 8  
|||||  
|  
  
RESULT 9  
Y79128  
ID Y79128 standard; Peptide: 8 AA.  
XX  
AC Y79128;  
XX  
DT 05-JUN-2000 (first entry)  
XX  
DE Peptide antagonist of zonulin.  
XX  
KW Zonulin; antagonist; zonula occludens toxin receptor;  
KW blood-brain barrier; antiinflammatory; cerebroprotective;  
KW neuroprotective; dermatological; antitumor; antiviral;  
KW antibacterial; cytostatic; anti-HIV; vulnerrary; antiallergic;  
KW hypotensive; immunosuppressive; antiparasitic; vasotropic;  
KW gastrointestinal inflammation; therapy.  
XX  
OS Synthetic.  
XX  
PN WO200007609-A1.  
XX  
PD 17-FEB-2000.  
XX  
PE 28-JUL-1999; 99WO-US16683.  
XX  
PR 03-AUG-1998; 98US-0127815.

XX	(UYMA-) UNIV MARYLAND BALTIMORE.
PA	Fasano A;
XX	
XX	WPI; 2000-205565/18.
DR	
XX	New peptide antagonist of zonulin useful as antiinflammatory agent for
PT	treating cerebral ischemia, stroke, cerebral edema, gastritis,
PT	shigellosis, viral gastroenteritis, meningitis, encephalomyelitis -
PS	Claim 1; Page 48; 69pp; English.
XX	
CC	This present sequence is that of a peptide antagonist of zonulin
CC	(2), one of 25 such peptides (see r79105-29) of the invention,
CC	which bind to a zonula occludens toxin (ZOT) receptor, yet do not
CC	physiologically modulate the opening of mammalian tight junctions
CC	(17). The peptide antagonists are based on a common motif of ZOT
CC	and human zonulins, which is believed to be critical for receptor
CC	binding. They can be prepared by chemical synthesis or by use of
CC	recombinant DNA techniques. The peptide antagonists are used as an
CC	antiinflammatory agents in the treatment of gastrointestinal
CC	inflammation, where they bind to the ZOT receptor in the intestine
CC	and yet does not physiologically modulate the opening of TJ in the
CC	intestine. Gastrointestinal inflammation conditions give rise to
CC	increased intestinal permeability and the peptide is useful for
CC	treating intestinal conditions that cause protein losing enteropathy
CC	caused by infection, e.g. Clostridium difficile infection,
CC	enterocolitis, shigellosis, viral gastroenteritis, parasite
CC	infestation, bacterial overgrowth, whipple's disease, diseases with
CC	mucosal erosion or ulcerations, e.g. gastritis, gastric cancer,
CC	collagenous colitis, inflammatory bowel disease, diseases marked by
CC	lymphatic obstruction, e.g. congenital intestinal lymphangiectasia,
CC	sarcoidosis lymphoma, mesenteric tuberculosis, and after surgical
CC	correction of congenital heart disease with Fontan's operation,
CC	mucosal diseases without ulceration, e.g. Menetrier's disease,
CC	coliac disease, eosinophilic gastroenteritis, and immune diseases,
CC	e.g. systemic lupus erythematosus or food allergies, primarily to
CC	milk.
SQ	Sequence 8 AA;
Query Match	82.9%; Score 34; DB 21; Length 8;
Best Local Similarity	87.5%; Pred. No. 3.2e+05;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;	
OY	1 GGVLVDG 8     1 ggglvqdg 8
Db	
RESULT 10	
W94489	
ID	W94489 standard; peptide; 20 AA.
AC	W94489;
DT	21-APR-1999 (first entry)
DE	Human foetal intestine zonulin N-terminal peptide.
XX	
KW	Zonulin; mammalian tight junction; zonula occludens toxin; ZOT;
KW	Vibrio cholerae; vacellae; cholera toxin; polyclonal antibody;
KW	Intestinal mucosa; nasal mucosa; blood brain barrier.
OS	Homo sapiens.
Key	Location/Qualifiers
Misc-difference 16	/note= "unspecified"
MO9852415-A1.	

[illegible]



XX WO200007609-A1.  
 PN 17-FEB-2000.  
 XX  
 XX 28-JUL-1999; 99WO-US16683.  
 XX 03-AUG-1998; 98US-0127815.  
 XX (UYMA-) UNIV MARYLAND BALTIMORE.  
 PA  
 PI Fasanio A;  
 XX WPI: 2000-205565/18.  
 DR  
 PT New peptide antagonist of zonulin useful as antiinflammatory agent for  
 PT treating cerebral ischemia, stroke, cerebral edema, gastritis,  
 PT shigellosis, viral gastroenteritis, meningitis, encephalomyelitis -  
 XX  
 PS Example 3; Fig 6; 69pp; English.  
 XX  
 CC The present sequence is that of the N-terminal region of foetal  
 CC human intestinal zonulin. The N-terminal sequences of human adult  
 CC and foetal zonulins (see Y79130-36) were compared with Vibrio cholerae  
 CC zonula occludens toxin (Zot) to identify a common motif thought  
 CC to be involved in receptor binding. Peptide antagonists (see  
 CC Y79105-29) based on this motif are useful as antiinflammatory  
 CC agents for treatment of gastrointestinal inflammation, and for  
 CC treatment of conditions associated with breakdown of the blood-brain  
 CC barrier.  
 CC  
 XX  
 SQ Sequence 20 AA:

Query Match 82.9%; Score 34; DB 21; Length 20;  
 Best Local Similarity 87.5%; Pred. No. 4.5;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GGVLVQDG 8  
 DB 8 ggvlvqpg 15

RESULT 12  
 RS2775  
 ID R52775 standard; Protein: 25 AA.  
 XX  
 AC R52775;  
 XX  
 DT 24-JAN-1995 (first entry)  
 XX  
 DE Murine Bre-3 immunoglobulin heavy chain variable domain N-terminus.  
 KW Immunoglobulin variable domain; primer; polymerase chain reaction;  
 KW chimeric antibody; human milk fat globule; Bre-3 VL-chain.  
 XX  
 OS Mus musculus.  
 XX  
 PN W09411508-A.  
 XX  
 PD 26-MAY-1994.  
 XX  
 PF 15-NOV-1993; 93WO-US11316.  
 XX  
 PR 13-NOV-1992; 92US-0977706.  
 PR 13-NOV-1992; 92US-0977707.  
 PR 28-SEP-1993; 93US-0128015.  
 XX  
 PA (CANC-) CANCER RES FUND CONTRA COSTA.  
 XX  
 DR WPI: 1994-183509/22.  
 XX  
 PT Chimeric human-murine polypeptide(s) specific for human mammary

PT fat globule antigen - for imaging, diagnosing and treating  
 PT neoplasia, with less undesirable immunogenic response  
 XX  
 PS Example 11; Page 32; 54pp; English.  
 XX  
 CC Primers J02, J03, J04 and VH1BACK (062740-062744) were all  
 CC used to prepare cDNAs that encode the Bre-3 mouse Ig variable  
 CC domains. The amplified V-regions lacked constant regions so as to  
 CC produce less immunogenic polypeptides. A hybrid polypeptide was  
 CC prepared using human constant regions with the murine V regions.  
 CC The chimeric polypeptide retained the binding affinity of Bre-3  
 CC for human milk fat globule. The amino acid sequences deduced from  
 CC the amplified VL and VH regions were compared to N-terminal  
 CC sequences which had been determined directly (i.e. R52774 and  
 CC R52775). The general agreement between the predicted and the  
 CC determined amino acid sequences that the cloned cDNAs encode Bre-3.  
 XX  
 SQ Sequence 25 AA:

Query Match 82.9%; Score 34; DB 15; Length 25;  
 Best Local Similarity 87.5%; Pred. No. 5.7;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GGVLVQDG 8  
 DB 8 ggvlvqpg 15

RESULT 13  
 RS2793  
 ID R52793 standard; Protein: 25 AA.  
 XX  
 AC R52793;  
 XX  
 DT 24-JAN-1995 (first entry)  
 XX  
 DE Murine Bre-3 immunoglobulin heavy chain variable domain N-terminus.  
 KW Immunoglobulin variable domain; primer; polymerase chain reaction;  
 KW chimeric antibody; human milk fat globule; Bre-3 VL-chain.  
 XX  
 OS Mus musculus.  
 XX  
 PN W09411509-A.  
 XX  
 PD 26-MAY-1994.  
 XX  
 PF 16-NOV-1993; 93WO-US11445.  
 XX  
 PR 16-NOV-1992; 92US-0977696.  
 PR 30-SEP-1993; 93US-0129930.  
 PR 08-OCT-1993; 93US-0134346.  
 XX  
 PA (CANC-) CANCER RES FUND CONTRA COSTA.  
 XX  
 DR WPI: 1994-183510/22.  
 XX  
 PT New analogue peptide(s) comprising antibody variable regions -  
 PT used to develop prods. for use in the detection, diagnosis,  
 PT therapy and prevention of neoplasms  
 XX  
 PS Example 12; Page 53; 54pp; English.  
 XX  
 CC Primers J02, J03, J04 and VH1BACK (062765-062769) were all  
 CC used to prepare cDNAs that encode the Bre-3 mouse Ig variable  
 CC domains. The amplified V-regions lacked constant regions so as to  
 CC produce less immunogenic polypeptides. A hybrid polypeptide was  
 CC prepared using human constant regions with the murine V regions.  
 CC The chimeric polypeptide retained the binding affinity of Bre-3  
 CC for human milk fat globule. The amino acid sequences deduced from  
 CC the amplified VL and VH regions were compared to N-terminal  
 CC sequences which had been determined directly (i.e. R52792 and

CC R52793). The general agreement between the predicted and the  
 CC determined amino acid sequences that the cloned cDNAs encode BRE-3.  
 XX

SO Sequence 25 AA:

Query Match 82.9%; Score 34; DB 15; Length 25;  
 Best Local Similarity 87.5%; Pred. No. 5.7;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GGVLYVDG 8  
 | | | | |  
 DB 8 ggvlyvpg 15

RESULT 14  
 Y79108  
 ID Y79108 standard; Peptide: 8 AA.  
 XX Y79108;  
 AC  
 XX 05-JUN-2000 (first entry)  
 DT  
 XX Peptide antagonist of zonulin.  
 DE  
 XX

XX Zonulin: antagonist; zonula occludens toxin receptor;  
 KW blood-brain barrier; antiinflammatory; cerebroprotective;  
 KW neuroprotective; dermatological; antitumor; antiviral;  
 KW antibacterial; cytostatic; anti-HIV; vulnery; antiallergic;  
 KW hypotensive; immunosuppressive; antiparasitic; vasotropic;  
 KW gastrointestinal inflammation; therapy.

OS Synthetic.

XX WO200007609-A1.

XX 17-FEB-2000.

XX 28-JUL-1999; 99WO-US16683.

XX 03-AUG-1998; 98US-0127815.

XX (UYMA-) UNIV MARYLAND BALTIMORE.

XX Fasano A;

XX WPI: 2000-205565/18.

XX New peptide antagonist of zonulin useful as antiinflammatory agent for  
 PT treating cerebral ischemia, stroke, cerebral edema, gastritis,  
 PT shigellosis, viral gastroenteritis, meningitis, encephalomyelitis  
 XX  
 PS Claim 1; Page 42; 69pp; English.

XX This present sequence is that of a peptide antagonist of zonulin  
 CC (2), one of 25 such peptides (see Y79105-29) of the invention,  
 CC which bind to a zonula occludens toxin (ZOT) receptor, yet do not  
 CC physiologically modulate the opening of mammalian tight junctions  
 CC (TJ). The peptide antagonists are based on a common motif of ZOT  
 CC and human zonulins, which is believed to be critical for receptor  
 CC binding. They can be prepared by chemical synthesis or by use of  
 CC recombinant DNA techniques. The peptide antagonists are used as an  
 CC antiinflammatory agents in the treatment of gastrointestinal  
 CC inflammation, where they bind to the ZOT receptor in the intestine  
 CC and yet does not physiologically modulate the opening of TJ in the  
 CC intestine. Gastrointestinal inflammation conditions give rise to  
 CC increased intestinal permeability and the peptide is useful for  
 CC treating intestinal conditions that cause protein losing enteropathy  
 CC caused by infection, e.g. Clostridium difficile infection.  
 CC enterocolitis, shigellosis, viral gastroenteritis, parasite  
 CC infection, bacterial overgrowth, whipple's disease, diseases with  
 CC mucosal erosion or ulcerations, e.g. gastritis, gastric cancer,  
 CC collagenous colitis, inflammatory bowel disease, diseases marked by

CC lymphatic obstruction, e.g. congenital intestinal lymphangiectasia,  
 CC sarcoidosis lymphoma, mesenteric tuberculosis, and after surgical  
 CC correction of congenital heart disease with Fontan's operation,  
 CC mucosal diseases without ulceration, e.g. Menetrier's disease,  
 CC coeliac disease, eosinophilic gastroenteritis, and immune diseases,  
 CC e.g. systemic lupus erythematosus or food allergies, primarily to  
 CC milk.  
 CC  
 XX  
 SO Sequence 8 AA:

Query Match 80.5%; Score 33; DB 21; Length 8;  
 Best Local Similarity 87.5%; Pred. No. 3.2e+05;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GGVLYVDG 8  
 | | | | |  
 DB 1 ggvlyvpg 8

RESULT 15  
 Y79122  
 ID Y79122 standard; Peptide: 8 AA.  
 XX Y79122;  
 AC  
 XX 05-JUN-2000 (first entry)  
 DT  
 XX Peptide antagonist of zonulin.  
 DE  
 XX

XX Zonulin: antagonist; zonula occludens toxin receptor;  
 KW blood-brain barrier; antiinflammatory; cerebroprotective;  
 KW neuroprotective; dermatological; antitumor; antiviral;  
 KW antibacterial; cytostatic; anti-HIV; vulnery; antiallergic;  
 KW hypotensive; immunosuppressive; antiparasitic; vasotropic;  
 KW gastrointestinal inflammation; therapy.

OS Synthetic.

XX WO200007609-A1.

XX 17-FEB-2000.

XX 28-JUL-1999; 99WO-US16683.

XX 03-AUG-1998; 98US-0127815.

XX (UYMA-) UNIV MARYLAND BALTIMORE.

XX Fasano A;

XX WPI: 2000-205565/18.

XX New peptide antagonist of zonulin useful as antiinflammatory agent for  
 PT treating cerebral ischemia, stroke, cerebral edema, gastritis,  
 PT shigellosis, viral gastroenteritis, meningitis, encephalomyelitis  
 XX  
 PS Claim 1; Page 46; 69pp; English.

XX This present sequence is that of a peptide antagonist of zonulin  
 CC (2), one of 25 such peptides (see Y79105-29) of the invention,  
 CC which bind to a zonula occludens toxin (ZOT) receptor, yet do not  
 CC physiologically modulate the opening of mammalian tight junctions  
 CC (TJ). The peptide antagonists are based on a common motif of ZOT  
 CC and human zonulins, which is believed to be critical for receptor  
 CC binding. They can be prepared by chemical synthesis or by use of  
 CC recombinant DNA techniques. The peptide antagonists are used as an  
 CC antiinflammatory agents in the treatment of gastrointestinal  
 CC inflammation, where they bind to the ZOT receptor in the intestine  
 CC and yet does not physiologically modulate the opening of TJ in the  
 CC intestine. Gastrointestinal inflammation conditions give rise to  
 CC increased intestinal permeability and the peptide is useful for  
 CC treating intestinal conditions that cause protein losing enteropathy

CC caused by infection, e.g. Clostridium difficile infection,  
CC enterocolitis, shigellosis, viral gastroenteritis, parasite  
CC infestation, bacterial overgrowth, whipple's disease, diseases with  
CC mucosal erosion or ulcerations, e.g. gastritis, gastric cancer,  
CC collagenous colitis, inflammatory bowel disease, diseases marked by  
CC lymphatic obstruction, e.g. congenital intestinal lymphangiectasia,  
CC sarcoidosis lymphoma, mesenteric tuberculosis, and after surgical  
CC correction of congenital heart disease with Fontan's operation,  
CC mucosal diseases without ulceration, e.g. Menetrier's disease,  
CC coeliac disease, eosinophilic gastroenteritis, and immune diseases,  
CC e.g. systemic lupus erythematosus or food allergies, primarily to  
CC milk.  
CC  
XX  
SQ Sequence 8 AA;

Query Match 80.5%; Score 33; DB 21; Length 8;  
Best Local Similarity 75.0%; Pred. No. 3.2e+05;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
QY 1 GGYLVQDG 8  
||:||||  
Db 1 gglcvqdg 8

Search completed: June 13, 2001, 14:14:39  
Job time: 380 sec

*This Page Blank (uspto)*

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

# OW protein - protein search, using sw model

Run on: June 13, 2001, 14:10:46 ; Search time 130.61 Seconds  
(without alignments)  
4.209 Million cell updates/sec

Title: PCT-US01-05825a-16

Perfect score: 41

Sequence: 1 GGVLYODG 8

## Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 198801 seqs, 68722935 residues

Total number of hits satisfying chosen parameters: 198801

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

## Database :

1: pir67:\*  
2: pir1:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	37	90.2	1542	2	T17459 polyprotein - toma
2	36	87.8	115	1	A46279 guanylin precursor
3	35	85.4	305	2	F69947 phage-related prot
4	35	85.4	537	2	F70597 hypothetical prote
5	35	85.4	606	2	A75554 ABC transporter, A
6	35	85.4	757	2	C32052 fixl protein - Rhl
7	35	85.4	1510	2	T33100 hypothetical prote
8	34	82.9	396	1	K4802 plasma protein 2 -
9	34	82.9	403	2	S42532 hypothetical prote
10	34	82.9	937	2	S78561 CS3 pilin synthesi
11	33	80.5	157	2	B70655 2-demethylmenaqu
12	33	80.5	213	2	A75471 probable nitroge
13	33	80.5	353	2	S61491 inclusion memb
14	33	80.5	368	2	T09401 1,3-beta-glucanase
15	33	80.5	370	2	T06552 glucan endo-1,3-be
16	33	80.5	418	2	B64924 hypothetical prote
17	33	80.5	897	2	E65127 hypothetical 97.1
18	32	78.0	134	2	A70315 AP4A hydrolase - A
19	32	78.0	234	2	B83065 conserved hypothet
20	32	78.0	242	2	H63334 glutamine transpor
21	32	78.0	258	2	S38459 replication protei
22	32	78.0	259	2	S30950 gene 2 protein - M
23	32	78.0	295	2	G83638 hypothetical prote
24	32	78.0	315	2	S27784 phenylethanolamine
25	32	78.0	319	2	S62191 5-methyltetrahydro
26	32	78.0	401	2	T50548 GRP cyclohydrolase
27	32	78.0	412	2	H70915 probable pgk prote
28	32	78.0	449	2	F83627 hypothetical prote
29	32	78.0	508	1	D7EBPH purH bifunctional

30	32	78.0	529	1	D7ECPH purH bifunctional
31	32	78.0	530	2	C82344 phosphoribosylamin
32	32	78.0	532	2	B64100 purH bifunctional
33	32	78.0	769	2	B83307 probable molybdopt
34	32	78.0	825	2	H64083 biotin sulfoxide r
35	32	78.0	1034	2	T22166 hypothetical prote
36	32	78.0	1302	2	B41249 multidrug resist
37	31	75.6	98	2	S25929 Ig heavy chain V r
38	31	75.6	120	2	S44111 Ig heavy chain V-D
39	31	75.6	151	2	A83333 probable deaminase
40	31	75.6	169	2	B70233 hypothetical prote
41	31	75.6	169	2	G75319 hypothetical prote
42	31	75.6	182	2	A64834 fibrital-like prot
43	31	75.6	193	2	D82271 yfnc protein VC086
44	31	75.6	196	2	F71525 hypothetical prote
45	31	75.6	196	2	H81681 mat protein TC0628

## ALIGNMENTS

RESULT 1  
T17459  
polyprotein - tomato  
C:Species: Lycopersicon esculentum (tomato)  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 11-May-2000  
C:Accession: T17459  
R:Parniske, M.; Jones, J.D.  
Proc. Natl. Acad. Sci. U.S.A. 96, 5850-5855, 1999  
A:Title: Recombination between diverged clusters of the tomato Cf-9 plant disease res  
A:Reference number: 218801; MUID:99254130  
A:Accession: T17459  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1542 <PAR>  
A:Cross-references: EMBL:AF119040; NID:g4235640; PID:g4235644; PIDN:AA13304.1

Query Match 90.2%; Score 37; DB 2; Length 1542;  
Best Local Similarity 87.5%; Pred. No. 3;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGVLYODG 8  
DB 920 GGVLYODG 927

RESULT 2  
A46279  
guanylin precursor [validated] - human  
C:Species: Homo sapiens (man)  
C:Date: 21-Sep-1993 #sequence\_revision 26-May-1995 #text\_change 08-Dec-2000  
R:Wiegand, R.C.; Kato, J.; Huang, W.J.; Gillett, N.; Henzel, W.; Goeddel, D.V.  
Proc. Natl. Acad. Sci. U.S.A. 89, 9089-9093, 1992  
A:Title: Precursor structure, expression, and tissue distribution of human guanylin.  
A:Reference number: A46279; MUID:93028409  
A:Accession: A46279  
A:Molecule type: mRNA  
A:Residues: 1-115 <DEL>  
A:Cross-references: GB:M95174; NID:g306823; PIDN:AA58625.1; PID:g306824  
A:Note: sequence extracted from NCBI backbone (NCBIN:115377; NCBI:115378)  
R:Wiegand, R.C.; Kato, J.; Huang, W.J.; Gillett, N.; Henzel, W.; Goeddel, D.V.  
FEBS Lett. 311, 150-154, 1992  
A:Title: Human guanylin: cDNA isolation, structure, and activity.  
A:Reference number: S29228; MUID:93011964  
A:Accession: S29228  
A:Molecule type: mRNA  
A:Residues: 1-115 <WIE>  
A:Cross-references: GB:M97496; NID:g183414; PIDN:AA35915.1; PID:g183415  
R:Kuhn, M.; Ralda, M.; Adernann, K.; Schulz-Knappe, P.; Gerzer, R.; Helm, J.M.; Forss  
FEBS Lett. 318, 205-209, 1993  
A:Title: The circulating bioactive form of human guanylin is a high molecular weight

A:Reference number: S29807; MUID:93178628  
 A:Accession: S29807  
 A:Molecule type: protein  
 A:Residues: 22-68 <KUN>  
 A:Experimental source: plasma  
 A:Note: amino-terminal sequencing of mature form and molecular weight of mature form by n of the same receptor.  
 C:Genetics:  
 A:Gene: GDB:GUC2  
 A:Cross-references: GDB:136460; OMIM:139392  
 A:Map position: 1p35-1p34  
 C:Superfamily: guanylin  
 C:Keywords: hormone; intestine  
 F:1-21/Domain: signal sequence #status predicted <SIG>  
 F:22-115/Product: guanylin #status experimental <MAT>

Query Match 67.8%; Score 36; DB 1; Length 115;  
 Best Local Similarity 67.5%; Pred. No. 3.3;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GGVLYVDG 8  
 ||| |||  
 Db 20 GGVLYVDG 27

## RESULT 3

F69947  
 Phage-related protein homolog yqbk - *Bacillus subtilis*

C:Species: *Bacillus subtilis*  
 C:Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 15-Oct-1999  
 C:Accession: F69947

R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berta, C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Chik, A.; Ehrlich, S.D.; Emerson, P.T.; Estlin, K.D.; Errington, J.; Fabret, C.; Ferrari, E.; Nature 390, 249-256, 1997  
 A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gallier, J.; Harwood, C.R.; Henaute, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.; Koester, P.; Konings, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois, Y.; M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelli, R.; Rieger, M.; Rivolta, C.; Roche, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanlon, A.; Schuster, S.; Schuster, R.; Scofield, F.; Sekiguchi, J.; Sekowska, A.; Serot, T.; Winters, P.; Wipet, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.; A:Authors: Yoshikawa, H.F.; Zimmstein, E.; Yoshikawa, H.; Danchin, A.  
 A:Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.  
 A:Reference number: A69580; MUID:98044033  
 A:Accession: F69947  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-305 <KUN>  
 A:Cross-references: GB:2991.7; GB:AL009126; NID:g2634966; PIDN:CAB14549.1; PID:el183837;  
 A:Experimental source: strain 168  
 C:Genetics:  
 A:Gene: yqbk

Query Match 85.4%; Score 35; DB 2; Length 305;  
 Best Local Similarity 75.0%; Pred. No. 15;  
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 GGVLYVDG 8  
 ||| |||  
 Db 271 GGVLYVDG 278

## RESULT 4

F70597

hypothetical protein RV3885c - *Mycobacterium tuberculosis* (strain H37RV)

C:Species: *Mycobacterium tuberculosis*  
 C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 22-Oct-1999

C:Accession: F70597  
 R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, J.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holtroyd, Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skellern, S.; Squares, S.; Nature 393, 537-544, 1998  
 A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrett, B.G.  
 A:Title: Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome.  
 A:Reference number: A70500; MUID:98295987  
 A:Accession: F70597  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-537 <COL>  
 A:Cross-references: GB:294121; GB:AL123456; NID:g3261736; PIDN:CAB08089.1; PID:el31229  
 A:Experimental source: strain H37RV  
 C:Genetics:  
 A:Gene: RV3885c

Query Match 85.4%; Score 35; DB 2; Length 537;  
 Best Local Similarity 87.5%; Pred. No. 27;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GGVLYVDG 8  
 ||| |||  
 Db 87 GGVLYVDG 94

## RESULT 5

A75554

ABC transporter, ATP-binding protein, MspA family - *Deinococcus radiodurans* (strain R)

C:Species: *Deinococcus radiodurans*  
 C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 17-Mar-2000  
 C:Accession: A75554

R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.; Smith, H.O.; Venter, J.C.; Fraser, C.M.  
 Science 286, 1371-1377, 1999  
 A:Title: Genome sequence of the radioresistant bacterium *Deinococcus radiodurans* R1.

A:Reference number: A75554; MUID:20036896  
 A:Accession: A75554  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-606 <MHT>  
 A:Cross-references: GB:AE001878; GB:AE000513; NID:g6457820; PIDN:AAF09749.1; PID:g645  
 A:Experimental source: strain R1  
 C:Genetics:  
 A:Gene: DR0163  
 A:Map position: 1  
 C:Superfamily: unassigned ATP-binding cassette proteins; ATP-binding cassette homolog

Query Match 85.4%; Score 35; DB 2; Length 606;  
 Best Local Similarity 75.0%; Pred. No. 31;  
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 GGVLYVDG 8  
 ||| |||  
 Db 561 GGVLYVDG 568

## RESULT 6

C32052

FixI protein - *Rhizobium meliloti*

C:Species: *Rhizobium meliloti*  
 C:Date: 13-Jul-1989 #sequence\_revision 13-Jul-1989 #text\_change 11-Jan-2000  
 C:Accession: C32052; S32847; S39994

R:Kahn, D.; David, M.; Domergue, O.; Daveran, M.L.; Ghal, J.; Hirsch, P.R.; Batut, J.  
 J. Bacteriol. 171, 929-939, 1989  
 A:Title: *Rhizobium meliloti* fixGHI sequence predicts involvement of a specific cation  
 A:Reference number: A32052; MUID:8912173  
 A:Accession: C32052  
 A:Status: preliminary  
 A:Molecule type: DNA

A:Residues: 1-757 <KAH>  
 A:Cross-references: EMBL:Z21854; NID:949403; PIDN:CAA79907.1; PID:949414  
 C:Superfamily: Enterococcus copper-transporting ATPase COPA; ATPase nucleotide-binding  
 C:Keywords: Phosphoprotein; transmembrane protein  
 F:162-95/Domain: ATPase transduction domain homology <ATP>  
 F:557-699/Domain: ATPase nucleotide-binding domain homology <ATN>

Query Match 85.4%; Score 35; DB 2; Length 757;  
 Best Local Similarity 75.0%; Pred. No. 39;  
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 GGVLPDQ 8  
 |||||  
 DB 432 GGVMVKDG 439

## RESULT 7

hypothetical protein H11E01.3 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 18-Feb-2000

C:Accession: T33100

R:Geisel, C.; Mamsley, P.  
 submitted to the EMBL Data Library, May 1998

A:Description: The sequence of C. elegans cosmid H11E01.

A:Reference number: 221282

A:Accession: T33100

A>Status: preliminary; translated from GB/EMBL/DDAY

A:Molecule type: DNA

A:Residues: 1-1510 <GEI>

A:Cross-references: EMBL:AF067222; PIDN:AC17017.1; GSPDB:GN00028; CESP:H11E01.3

A:Experimental source: strain Bristol N2; clone H11E01

C:Genetics:

A:Gene: CESP:H11E01.3

A:Map position: X

A:introns: 73/2; 120/1; 178/3; 227/2; 262/1; 300/2; 345/1; 1430/3; 1468/2

Query Match 85.4%; Score 35; DB 2; Length 1510;  
 Best Local Similarity 75.0%; Pred. No. 82;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 GGVLPDQ 8  
 |||||  
 DB 1495 GGILVDDG 1502

## RESULT 8

plasma protein Z - bovine

C:Species: Bos primigenius taurus (cattle)

C:Date: 27-Nov-1985 #sequence\_revision 27-Nov-1985 #text\_change 16-Jul-1999

C:Accession: A22171; A00926

R:Hoelrup, P.; Jensen, M.S.; Petersen, T.E.  
 FEBS Lett. 184, 333-338, 1985

A:Title: Amino acid sequence of bovine protein Z: a vitamin K-dependent serine protease

A:Reference number: A22171; MUID:85204370

A:Accession: A22171

A:Molecule type: protein

A:Residues: 1-396 <HOE>

C:Comment: Protein Z is a single-chain plasma glycoprotein of unknown function. Although

and has no enzymatic activity.

C:Superfamily: coagulation factor X; EGF homology; Gla domain homology; trypsin homology

C:Keywords: beta-hydroxyaspartic acid; calcium binding; carboxylutamic acid; glycoprote

F:1-46/Domain: calcium binding #status predicted <CAB>

F:51-82/Domain: Gla domain homology (fragment) <GLA>

F:88-135/Domain: EGF homology <EG1>

F:143-352/Domain: trypsin homology <TRY>

F:7,8,11,15,17,20,21,26,27,30,33,36,40/Modified site: gamma-carboxyglutamic acid (Glu) #

F:59,191,289/binding site: carbohydrate (Asn) (covalent) #status experimental

F:64/Modified site: erythro-beta-hydroxyaspartic acid (Asp) #status predicted

F:388/Binding site: carbohydrate (Thr) (covalent) #status experimental

Query Match 82.9%; Score 34; DB 1; Length 396;  
 Best Local Similarity 85.7%; Pred. No. 32;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GGVLPDQ 7  
 |||||  
 DB 170 GGVLDD 176

## RESULT 9

hypothetical protein 1 - Synecchococcus sp.

C:Species: Synecchococcus sp.

C:Date: 19-Mar-1997 #sequence\_revision 19-Mar-1997 #text\_change 05-Mar-1999

C:Accession: S42532

R:Jones, M.C.; Jenkins, J.M.; Smith, A.G.; Howe, C.J.  
 Plant Mol. Biol. 24, 435-448, 1994

A:Title: Cloning and characterisation of genes for tetrapyrrole biosynthesis from the

A:Reference number: S42531; MUID:94169298

A:Accession: S42532

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-403 <JON>

A>Note: the source is designated as Anacystis nidulans R2

C:Superfamily: Synecchococcus hypothetical protein 1; tetrairicopeptide repeat homology

F:283-316/Domain: tetrairicopeptide repeat homology <TT1>

F:317-350/Domain: tetrairicopeptide repeat homology <TT2>

F:351-384/Domain: tetrairicopeptide repeat homology <TT3>

Query Match 82.9%; Score 34; DB 2; Length 403;  
 Best Local Similarity 75.0%; Pred. No. 32;  
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 GGVLPDQ 8  
 |||||  
 DB 215 GGLVDDG 222

## RESULT 10

CS3 pilin synthesis protein, 104K - Escherichia coli

N:Contains: CS3 pilin synthesis protein, 104K; CS3 pilin synthesis protein, 20K; CS3

C:Species: Escherichia coli

C:Date: 13-Sep-1998 #sequence\_revision 13-Sep-1998 #text\_change 21-Jul-2000

C:Accession: S78561; S07900; S07901; S07902; S07903

R:Jalajakumari, M.B.; Thomas, C.J.; Halter, R.; Manning, P.A.  
 Mol. Microbiol. 3, 1685-1695, 1989

A:Title: Genes for biosynthesis and assembly of CS3 pilin of CFA/II enterotoxigenic Es

A:Reference number: S07899; MUID:90158116

A:Accession: S78561

A:Molecule type: DNA

A:Residues: 1-937 <JAL>

A:Cross-references: EMBL:X16944

A:Experimental source: strain PB176

A:Accession: S07900

A:Molecule type: DNA

A:Residues: 181-753 <JAW>

A:Cross-references: EMBL:X16944; NID:941155; PIDN:CAA34816.1; PID:941157

A:Experimental source: strain PB176

A:Accession: S07901

A:Molecule type: DNA

A:Residues: 317-753 <JAP>

A:Cross-references: EMBL:X16944; NID:941155; PIDN:CAA34817.1; PID:941158

A:Experimental source: strain PB176

A:Accession: S07902

A:Molecule type: DNA

A:Residues: 451-753 <JAA>

A:Cross-references: EMBL:X16944; NID:941155; PIDN:CAA34818.1; PID:941159

A:Experimental source: strain PB176  
 A:Accession: S07903  
 A:Molecule type: DNA  
 A:Residues: 572-753 <JAZ>  
 A:Cross-references: EMBL:X16944; NID:g41155; PIDN:CAA34819.1; PID:g41160  
 A:Experimental source: strain PB176  
 C:Function:  
 A:Description: Involved in the biosynthesis of CS3 pili  
 F:1-937/Product: CS3 pili synthesis protein, 104k #status predicted <MA1>  
 F:181-753/Product: CS3 pili synthesis protein, 63k #status predicted <MA2>  
 F:317-753/Product: CS3 pili synthesis protein, 48k #status predicted <MA3>  
 F:451-753/Product: CS3 pili synthesis protein, 33k #status predicted <MA4>  
 F:572-753/Product: CS3 pili synthesis protein, 20k #status predicted <MA5>

Query Match 82.9%; Score 34; DB 2; Length 937;  
 Best Local Similarity 75.0%; Pred. No. 79;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 GGVLYVDG 8  
 ||:||||  
 Db 553 GGVLYVDG 560

RESULT 11  
 B70655  
 2-demethylmenaquinone 2-C-methyltransferase (EC 2.1.1.-) meng [similarity] - Mycobacteri  
 C:Species: Mycobacterium tuberculosis  
 C>Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 02-Sep-2000  
 C:Accession: B70655  
 R:Cole, S.T.; Broesch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.  
 ; Connor, R.; Davies, R.; Davlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holtroyd, S.  
 ; Rajendram, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
 ; Nature 393: 537-544, 1998  
 A:Authors: Sgares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
 A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
 A:Reference number: A70500; MUID:98295987  
 A:Accession: B70655  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-157 <COL>  
 A:Cross-references: GB:28384; GB:AL123456; NID:g3261687; PIDN:CAB06213.1; PID:e301444;  
 A:Experimental source: strain H37Rv  
 C:Genetics:  
 A:Gene: meng  
 C:Keywords: methyltransferase

Query Match 80.5%; Score 33; DB 2; Length 157;  
 Best Local Similarity 87.5%; Pred. No. 19;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GGVLYVDG 8  
 ||:||||  
 Db 61 GGVLYVDG 68

RESULT 12  
 A75471  
 probable nitrogen regulator - Deinococcus radiodurans (strain R1)  
 C:Species: Deinococcus radiodurans  
 C>Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 31-Mar-2000  
 C:Accession: A75471  
 R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;  
 ; M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Usterbach, T.; Zalewski, C.; Ma  
 Science 286: 1571-1577, 1999  
 A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.  
 A:Reference number: A75250; MUID:20036896  
 A:Accession: A75471  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-213 <WH1>

A:Cross-references: GB:AE001937; GB:AE000513; NID:g6458542; PIDN:AAF10412.1; PID:g645  
 A:Experimental source: strain R1  
 C:Genetics:  
 A:Gene: DR0834  
 A:Map position: 1  
 C:Superfamily: regulatory protein fmr; CAMP receptor protein cyclic nucleotide-binding

Query Match 80.5%; Score 33; DB 2; Length 213;  
 Best Local Similarity 75.0%; Pred. No. 26;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 GGVLYVDG 8  
 ||:||||  
 Db 121 GGVLYVDG 128

RESULT 13  
 S61491  
 inclusion membrane protein A - Chlamydia psittaci  
 C:Species: Chlamydia psittaci, Chlamydia psittaci  
 C>Date: 27-Apr-1996 #sequence\_revision 13-Mar-1997 #text\_change 31-Mar-2000  
 C:Accession: S61491  
 R:Rockey, D.D.; Heinzen, R.A.; Hackstadt, T.  
 Mol. Microbiol. 15: 617-626, 1995  
 A:Title: Cloning and characterization of a Chlamydia psittaci gene coding for a prote  
 A:Reference number: S61491; MUID:95302975  
 A:Accession: S61491  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-355 <ROC>  
 A:Cross-references: EMBL:L35036; NID:g516598; PIDN:AAC41443.1; PID:g516599

Query Match 80.5%; Score 33; DB 2; Length 355;  
 Best Local Similarity 85.7%; Pred. No. 45;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 GGVLYVDG 8  
 ||:||||  
 Db 318 GGVLYVDG 324

RESULT 14  
 T09401  
 1,3-beta-glucanase (EC 3.2.1.-), acidic - alfalfa  
 C:Species: Medicago sativa (alfalfa)  
 C>Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 13-Aug-1999  
 C:Accession: T09401  
 R:Clouse, J.A.  
 submitted to the EMBL Data Library, May 1995  
 A:Reference number: Z16664  
 A:Accession: T09401  
 A:Status: translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-368 <CLO>  
 A:Cross-references: EMBL:U27179; NID:9862930; PID:9862931  
 A:Experimental source: subspecies sativa, cultivar Apollo  
 C:Function:  
 A:Pathway: polysaccharide degradation  
 C:Superfamily: beta-1,3-glucanase  
 C:Keywords: glycosidase; hydrolase; polysaccharide degradation

Query Match 80.5%; Score 33; DB 2; Length 368;  
 Best Local Similarity 85.7%; Pred. No. 47;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 GGVLYVDG 8  
 ||:||||  
 Db 234 GGVLYVDG 240



## RESULT 15

glucan endo-1,3-beta-D-glucosidase (EC 3.2.1.39) - garden pea  
T06552  
N:Alternate names: beta-1,3-glucanase  
C:Species: Pisum sativum (garden pea)  
C:Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 18-Aug-2000  
C:Accession: T06552; S28430  
R:Chang, M.M.; Culley, D.E.; Hadwiger, L.A.  
Plant Physiol. 101, 1121-1122, 1993  
A:Title: Nucleotide sequence of a pea (Pisumsativum L.) beta-1,3-glucanase gene.  
A:Reference number: 215757; MUID:94143473  
A:Accession: T06552  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-370 <CHAI>  
A:Cross-references: EMBL:L02212; NID:g309669; PIDN:AAA33648.1; PID:g169047  
R:Chang, M.M.; Hadwiger, L.A.; Horowitz, D.  
Plant Mol. Biol. 20, 609-618, 1992  
A:Title: Molecular characterization of a pea beta-1,3-glucanase induced by Fusarium sole  
A:Reference number: S28430; MUID:93081722  
A:Accession: S28430  
A:Status: not compared with conceptual translation  
A:Molecule type: mRNA  
A:Residues: 32-370 <CHAI>  
A:Cross-references: GB:S51479; NID:g261211; PIDN:AAB24398.1; PID:g261212  
C:Genetics:  
A:Introns: 31/1  
C:Superfamily: beta-1,3-glucanase  
C:Keywords: glycosidase; hydrolase; polysaccharide degradation

Query Match 80.5%; Score 33; DB 2; Length 370;  
Best Local Similarity 85.7%; Pred. NO. 47;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 2 GVLVQDG 8  
||:||||  
Db 233 GVLVQDG 239

Search completed: June 13, 2001, 14:10:46  
Job time: 147 sec

*This Page Blank (uspto)*

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 13, 2001, 14:21:46 ; Search time 74.44 Seconds  
(without alignments)  
3.681 Million cell updates/sec

Title: PCT-US01-05825A-16

Perfect score: 41

Sequence: 1 GGVLVQDG 8

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : swissprot\_39:\*

Prod. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	36	87.8	115	1	GUAN_HUMAN
2	35	85.4	305	1	YOBK_BACSU
3	35	85.4	757	1	FIXI_RHIME
4	34	82.9	396	1	PRIZ_BOVIN
5	34	82.9	403	1	YCOA_SYNP7
6	34	82.9	397	1	CS32_ECOLI
7	33	80.5	370	1	E13B_PEA
8	33	80.5	396	1	A2BP_MOUSE
9	33	80.5	397	1	A2BP_HUMAN
10	33	80.5	897	1	CHTA_ECOLI
11	32	78.0	109	1	GUAN_PIG
12	32	78.0	258	1	RF42_CRIFA
13	32	78.0	259	1	VG02_BPMU5
14	32	78.0	273	1	YKA2_CAEEL
15	32	78.0	309	1	METE_METTM
16	32	78.0	401	1	GCH2_ACTPL
17	32	78.0	412	1	PKR_MYCTU
18	32	78.0	508	1	PUR9_SALTY
19	32	78.0	529	1	PUR9_ECOLI
20	32	78.0	532	1	PUR9_HAEIN
21	32	78.0	825	1	BISC_HAEIN
22	32	78.0	1302	1	MOR5_DROME
23	31	75.6	182	1	YCR0_ECOLI
24	31	75.6	204	1	RIB3_SCHPO
25	31	75.6	346	1	GUAC_ECOLI
26	31	75.6	512	1	PUR9_BACSU
27	31	75.6	1182	1	PRC2_MOUSE
28	31	75.6	1203	1	PRC2_HUMAN
29	31	75.6	1656	1	ATC8_YEAST
30	30	73.2	1569	1	YPOA_ECOLI
31	30	73.2	1107	1	GUAN_CAVPO
32	30	73.2	115	1	GUAN_RAT
33	30	73.2	116	1	GUAN_MOUSE

34	30	73.2	116	1	HV30_HUMAN	P01778 homo sapien
35	30	73.2	247	1	YCKI_BACSU	P39456 bacillus su
36	30	73.2	248	1	GRAE_MOUSE	P08884 mus musculu
37	30	73.2	248	1	GRAE_MOUSE	P08883 mus musculu
38	30	73.2	248	1	GRAE_MOUSE	P13366 mus musculu
39	30	73.2	274	1	DAPD_ACTPL	P41396 actinobacill
40	30	73.2	275	1	DAPD_HAEIN	P45284 haemophilus
41	30	73.2	347	1	MKL_MYCTE	P30769 mycobacteri
42	30	73.2	359	1	MKL_MYCTU	P06784 mycobacteri
43	30	73.2	362	1	FIBP_ADE09	P36846 human adeo
44	30	73.2	463	1	YD14_SCHPO	O92342 schizosacch
45	30	73.2	515	1	TLC1_CHLPN	O928j2 chlamydia p

#### ALIGNMENTS

RESULT 1  
ID GUAN\_HUMAN STANDARD: PRT: 115 AA.  
AC 002747;  
DT 01-JUL-1993 (Rel. 26, Created)  
DT 01-JUL-1993 (Rel. 26, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE GUANYLIN PRECURSOR (GUANYLATE CYCLASE ACTIVATOR 2A).  
GN GUCA2A OR GUCA2.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Ductenum;  
RX MEDLINE=93011964; PubMed=1327879;  
RA Wiegand R.C., Kato J., Huang M.D., Fox K.F., Kachur J.F.,  
RA Currie M.G.;  
RT "Human guanylin: cDNA isolation, structure, and activity.";  
RL FEBS Lett. 311:150-154(1992).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Ileum;  
RX MEDLINE=93028409; PubMed=1409606;  
RA de Sauvage F.J., Keshav S., Kuang W.J., Gillett N., Henzel W.,  
RA Goeddel D.V.;  
RT "Precursor structure, expression, and tissue distribution of human  
guanylin.";  
RL Proc. Natl. Acad. Sci. U.S.A. 89:9089-9093(1992).  
RN [3]  
RP SEQUENCE OF 22-68.  
RX MEDLINE=93178628; PubMed=8095028;  
RA Kuhn M., Raide M., Adermann K., Schulz-Knappe P., Gerzer R.,  
RA Heim J.-M., Forssmann W.-G.;  
RT "The circulating bioactive form of human guanylin is a high molecular  
weight peptide (10.3 kDa).";  
RL FEBS Lett. 318:205-209(1993).  
RN [4]  
RP STRUCTURE BY NMR OF 101-115.  
RX MEDLINE=95034794; PubMed=7947768;  
RA Skelton N.J., Garcia K.C., Goeddel D.V., Quan C., Burnier J.P.;  
RT "Determination of the solution structure of the peptide hormone  
guanylin: observation of a novel form of topological  
stereoisomerism.";  
RL Biochemistry 33:13581-13592(1994).  
CC -I- FUNCTION: ENDOGENOUS ACTIVATOR OF INTESTINAL GUANYLATE CYCLASE.  
CC IT STIMULATES THIS ENZYME THROUGH THE SAME RECEPTOR BINDING REGION  
CC AS THE HEAT-STABLE ENTEROTOXINS.  
CC -I- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN ILEUM AND COLON.  
CC -I- SIMILARITY: BELONGS TO THE GUANYLIN FAMILY. ALSO SIMILAR TO HEAT-  
CC STABLE ENTEROTOXINS.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: M97496; AA35915.1; -  
 DR EMBL: M95174; AA58625.1; -  
 DR PIR: A46279; A46279.  
 DR PIR: S29228; S29228.  
 DR PDB: 1GNA; 30-SEP-94.  
 DR PDB: 1GNB; 30-SEP-94.  
 DR MIM: 139392; -  
 DR InterPro: IPR000879; -  
 DR Pfam: PF02058; Guanylin; 1.  
 DR PRINTS: PRO0774; GUANYLIN.  
 KW Signal; 3D-structure.  
 FT SIGNAL 1 21  
 FT CHAIN 22 115 HMW-GUANYLIN.  
 FT PEPTIDE 101 115  
 FT DISULFID 104 112 GUANYLIN.  
 FT DISULFID 107 115  
 FT SEQUENCE 115 AA; 12148 MW; 664AC03BAFEC26FA CRC64;

Query Match Best Local Similarity 87.8%; Score 36; DB 1; Length 115;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 GGVLVODG 8  
 |||||  
 Db 20 GGVTVODG 27

RESULT 2  
 YOBK\_BACSU STANDARD; PRT; 305 AA.  
 AC PA5927;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 01-OCT-1996 (Rel. 34, Last annotation update)  
 DE HYPOTHEICAL 32.7 KDA PROTEIN IN SPOILIC-CWLA INTERGENIC REGION.  
 GN YOBK.  
 OS Bacillus subtilis.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
 CC Bacillus/Staphylococcus group; Bacillus.  
 OX NCBI\_TaxID=1423;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168 / JH642;  
 RX MEDLINE=95219086; PubMed=7704261;  
 RT Takemaru K.-I., Mizuno M., Sato T., Takeuchi M., Kobayashi Y.;  
 RT "Complete nucleotide sequence of a skin element excised by DNA  
 RT rearrangement during sporulation in Bacillus subtilis.";   
 RL Microbiology 141:323-327(1995).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168 / JH642;  
 RA Kobayashi Y., Mizuno M., Masuda S., Takemaru K., Hosono S.,  
 RA Sato T., Takeuchi M.,  
 RL Submitted (May-1996) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP IDENTIFICATION.  
 RX MEDLINE=96084975; PubMed=7489895;  
 RA Medigue C., Moszer I., Viari A., Danchin A.;  
 RT "Analysis of a Bacillus subtilis genome fragment using a co-operative  
 RT computer system prototype.";   
 RL Gene 165:637-651(1995).  
 CC -1- SIMILARITY: STRONG, TO B.SUBTILIS XKDK.  
 CC -1- CAUTION: REF.1 SEQUENCE DIFFERS FROM THAT SHOWN FROM POSITION 179  
 CC ONWARD AND IS SHORTER (226 AA) DUE TO A FRAMESHIFT.  
 CC -1- This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: D32216; BA06943.1; ALT\_FRAME.  
 DR EMBL: D84432; BA12406.1; -  
 DR EMBL: Z99117; CAB14549.1; -  
 DR Subtilisin; BG1282; YgbK.  
 KW Hypothetical protein.  
 SO SEQUENCE 305 AA; 32667 MW; FE12A8987867B1EF CRC64;

Query Match Best Local Similarity 85.4%; Score 35; DB 1; Length 305;  
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 GGVLVODG 8  
 |||||  
 Db 271 GGVLENDG 278

RESULT 3  
 FIXI\_RHIME STANDARD; PRT; 757 AA.  
 ID FIXI\_RHIME  
 AC P18398;  
 DT 01-NOV-1990 (Rel. 16, Created)  
 DT 01-NOV-1990 (Rel. 16, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE NITROGEN FIXATION PROTEIN FIXI (EI-E2 TYPE CATION ATPASE FIXI)  
 DE (EC 3.6.1.-).  
 GN FIXI.  
 OS Rhizobium meliloti (Sinorhizobium meliloti).  
 OC Plasmid pSymA (megaplasmid 1).  
 CC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
 OC Rhizobiaceae; Sinorhizobium.  
 OX NCBI\_TaxID=382;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=RCR2011 / Su47;  
 RX MEDLINE=89123173; PubMed=2536685;  
 RA Kahn D., David M., Domergue O., Daveran M.-L., Ghal J., Hirsch P.R.,  
 RA Batut J.;  
 RT "Rhizobium meliloti fixGHI sequence predicts involvement of a  
 RT specific cation pump in symbiotic nitrogen fixation.";   
 RL J. Bacteriol. 171:929-939(1989).  
 CC -1- FUNCTION: FIXI IS A PUMP OF A SPECIFIC CATION INVOLVED IN  
 CC SYMBIOTIC NITROGEN FIXATION. THE FOUR PROTEINS FIXG, FIXH, FIXI,  
 CC AND FIXJ MAY PARTICIPATE IN A MEMBRANE-BOUND COMPLEX COUPLING  
 CC THE FIXI CATION PUMP WITH A REDOX PROCESS CATALYZED BY FIXG.  
 CC -1- CATALYTIC ACTIVITY: ATP + H(2)O = ADP + ORTHOPHOSPHATE.  
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.  
 CC -1- SIMILARITY: BELONGS TO THE CATION TRANSPORT ATPASES FAMILY  
 CC (EI-E2 ATPASES). SUBFAMILY IB.  
 CC -1- SIMILARITY: CONTAINS 1 HEAVY-METAL-ASSOCIATED (HMA) DOMAIN.  
 CC -1- This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: M24144; AA26273.1; -  
 DR EMBL: Z21854; CAA79807.1; -  
 DR PIR: C32052; C32052.  
 DR PIR: S39994; S39994.  
 DR PIR: S32847; S32847.  
 DR HSSP: Q27546; IMAS.  
 DR InterPro: IPR001757; -  
 DR InterPro: IPR001934; -

DR Pfam: PF00122; E1-E2\_ATPase; 1.  
DR Pfam: PF00403; HMA; 1.  
DR PROSITE: PS00154; ATPASE\_E1\_E2; 1.  
DR PROSITE: PS01047; HMA; 1.  
KW Hydrolyase; Transmembrane; Phosphorylation; ATP-binding; Metal-binding;  
KM Nitrogen fixation; Plasmid.  
FT DOMAIN 1 123 CYTOPLASMIC (POTENTIAL).  
FT TRANSSEM 124 144 EXTRACELLULAR (POTENTIAL).  
FT DOMAIN 145 154 EXTRACELLULAR (POTENTIAL).  
FT TRANSSEM 155 175 POTENTIAL.  
FT DOMAIN 176 209 CYTOPLASMIC (POTENTIAL).  
FT TRANSSEM 210 230 POTENTIAL.  
FT DOMAIN 231 374 EXTRACELLULAR (POTENTIAL).  
FT TRANSSEM 375 395 POTENTIAL.  
FT DOMAIN 396 397 CYTOPLASMIC (POTENTIAL).  
FT TRANSSEM 398 418 POTENTIAL.  
FT DOMAIN 419 590 EXTRACELLULAR (POTENTIAL).  
FT TRANSSEM 591 611 POTENTIAL.  
FT DOMAIN 612 689 CYTOPLASMIC (POTENTIAL).  
FT TRANSSEM 690 710 POTENTIAL.  
FT DOMAIN 711 711 EXTRACELLULAR (POTENTIAL).  
FT TRANSSEM 712 732 POTENTIAL.  
FT DOMAIN 733 757 CYTOPLASMIC (POTENTIAL).  
FT DOMAIN 43 72 HMA.  
FT METAL 48 48 POTENTIAL.  
FT METAL 51 51 POTENTIAL.  
FT MOD\_RES 454 454 PHOSPHORYLATION (PROBABLE).  
FT SEQUENCE 757 AA; 79559 MW; DE2E5C6249254AA5 CRC64;  
  
Query Match 85.4%; Score 35; DB 1; Length 757;  
Best Local Similarity 75.0%; Pred. No. 21;  
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
  
OY 1 GGVLYVDG 8  
Db 432 GGVWVKDS 439  
|||:|:|  
|:|:|:|  
  
RESULT 4  
PRTZ\_BOVIN STANDARD; PRT; 396 AA.  
AC P00744:  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 21-JUL-1986 (Rel. 01, Last sequence update)  
DT 15-JUL-1998 (Rel. 36, Last annotation update)  
DE VITAMIN K-DEPENDENT PROTEIN Z.  
GN PROZ.  
OS Bos taurus (Bovine).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
OC Bovidae; Bovinae; Bos.  
OX NCBI\_TaxID=9913;  
RN [1]  
RP SEQUENCE  
RA MEDLINE=85204370; PubMed=3886670;  
RA Hoejrup P., Jensen M.S., Petersen T.E.;  
RT "Amino acid sequence of bovine protein Z: a vitamin K-dependent  
RT serine protease homolog.";  
RL FEBS Lett. 184:333-336(1985).  
RN [2]  
RP STRUCTURE OF CARBOHYDRATE ON SER-53.  
RA MEDLINE=90062160; PubMed=2511201;  
RA Nishimura H., Kawabata S., Kistiel W., Hase S., Ikenaka T., Takao T.,  
RA Shimonishi Y., Iwanaga S.;  
RT "Identification of a disaccharide (Xyl-Glc) and a trisaccharide  
RT (Xyl2-Glc) O-glycosidically linked to a serine residue in the first  
RT epidermal growth factor-like domain of human factors VII and IX and  
RT protein Z and bovine protein Z.";  
RL J. Biol. Chem. 264:20320-20325(1989).  
RN [3]  
RP STRUCTURE OF CARBOHYDRATE ON SER-53.  
RA MEDLINE=91344709; PubMed=2129367;

RA Iwanaga S., Nishimura H., Kawabata S., Kistiel W., Hase S., Ikenaka T.;  
RT "A new trisaccharide sugar chain linked to a serine residue in the  
RT first EGF-like domain of clotting factors VII and IX and protein Z.";  
RL Adv. Exp. Med. Biol. 281:121-131(1990).  
CC -1- FUNCTION: SINGLE-CHAIN PLASMA GLYCOPROTEIN OF UNKNOWN FUNCTION.  
CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR.  
CC -1- TISSUE SPECIFICITY: PLASMA.  
CC -1- SIMILARITY: ALTHOUGH HOMOLOGOUS WITH THE VITAMIN K-DEPENDENT  
CC CLOTTING FACTORS, IT HAS LOST TWO OF THE ESSENTIAL CATALYTIC  
CC RESIDUES AND HAS NO ENZYMIC ACTIVITY.  
CC -1- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.  
DR PIR: A22171; KXBOZ.  
DR HSSP: P00740; ICFH.  
DR MEROPS: S01.979;  
DR Glycosylated: P00744; -;  
DR InterPro: IPR00152; -;  
DR InterPro: IPR000294; -;  
DR InterPro: IPR000561; -;  
DR InterPro: IPR001254; -;  
DR InterPro: IPR002383; -;  
DR Pfam: PF00006; EGF; 2.  
DR Pfam: PF00594; gla; 1.  
DR Pfam: PF00089; trypsin; 1.  
DR PRINTS: PRO0001; GLABLOD.  
DR PROSITE: PS00010; ASX\_HYDROXYL; 1.  
DR PROSITE: PS00011; GLU\_CARBOXYLATION; 1.  
DR PROSITE: PS00022; EGF\_1; 1.  
DR PROSITE: PS01186; EGF\_2; 1.  
KW Plasma; Glycoprotein; Gamma-carboxyglutamic acid; Hydroxylation;  
KM Calcium; Serine protease homolog; Vitamin K; EGF-like domain.  
FT DOMAIN 47 83  
FT DOMAIN 85 126  
FT DOMAIN 143 396  
FT MOD\_RES 7 7  
FT MOD\_RES 8 8  
FT MOD\_RES 11 11  
FT MOD\_RES 15 15  
FT MOD\_RES 17 17  
FT MOD\_RES 20 20  
FT MOD\_RES 21 21  
FT MOD\_RES 26 26  
FT MOD\_RES 27 27  
FT MOD\_RES 30 30  
FT MOD\_RES 33 33  
FT MOD\_RES 36 36  
FT MOD\_RES 40 40  
FT MOD\_RES 64 64  
FT DISULFID 51 62  
FT DISULFID 56 71  
FT DISULFID 73 82  
FT DISULFID 89 101  
FT DISULFID 97 110  
FT DISULFID 112 125  
FT CARBOHYD 53 53  
FT CARBOHYD 59 59  
FT CARBOHYD 191 191  
FT CARBOHYD 289 289  
FT CARBOHYD 388 388  
SO SEQUENCE 396 AA; 43112 MW; 04C5D7A35849B116 CRC64;  
  
Query Match 82.9%; Score 34; DB 1; Length 396;  
Best Local Similarity 85.7%; Pred. No. 17;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
  
OY 1 GGVLYVD 7  
Db 170 GGVLYVD 176  
|||:|:|  
|:|:|:|  
  
RESULT 5  
YCOA\_SYNP7 STANDARD; PRT; 403 AA.  
ID YCOA\_SYNP7

AC P42460;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DE 15-JUL-1998 (Rel. 36, Last annotation update)  
 DE HYPOTHEORETICAL 45.0 KDA PROTEIN IN CORA 5 REGION.  
 OS Synecococcus sp. (Strain PCC 7942) (Anacystis nidulans R2).  
 CC Bacteria; Cyanobacteria; Chroococcales; Synecococcus.  
 CC NCBI\_TaxID=1140;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=94169298; PubMed=8123787;  
 RA Jones M.C., Jenkins J.M., Smith A.G., Howe C.J.;  
 RT "Cloning and characterization of genes for tetrapyrrole biosynthesis  
 from the cyanobacterium Anacystis nidulans R2.";  
 RL Plant Mol. Biol. 24:431-448(1994).  
 CC -1- SIMILARITY: IN THE N-TERMINAL, STRONG TO H. INFLUENZAE HI0653.  
 CC -1- SIMILARITY: CONTAINS 5 TPR REPEATS.  
 CC -1- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-14 IS THE INITIATOR.  
 CC -----  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: X70966; CAA50301.1; -  
 DR InterPro: IPR001173; -  
 DR DR InterPro: IPR001440; -  
 DR Pfam: PF00535; Glycosyltransferase\_2; 1.  
 DR Pfam: PF00515; TPR; 1  
 KM Hypothetical protein; Repeat: TPR repeat.  
 FT REPEAT 208 243 TPR 1.  
 FT REPEAT 244 282 TPR 2.  
 FT REPEAT 283 316 TPR 3.  
 FT REPEAT 317 350 TPR 4.  
 FT REPEAT 351 387 TPR 5.  
 FT REPEAT  
 SQ SEQUENCE 403 AA; 44998 MW; 4C0AB39B54988A13 CRC64;  
 Query Match 82.9%; Score 34; DB 1; Length 403;  
 Best Local Similarity 75.0%; Pred. No. 17;  
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
 Oy 1 GGVLYODG 8  
 Db 215 GGLVYDGG 222  
 RESULT 6  
 CS32\_ECOLI STANDARD: PRT; 937 AA.  
 ID CS32\_ECOLI  
 AC P15484; P15485; P15486; P15487;  
 DT 01-APR-1990 (Rel. 14, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE OUTER MEMBRANE USHER PROTEIN CS3-2 PRECURSOR (CS3 PILI SYNTHESIS 104  
 DE KDA PROTEIN).  
 OS Escherichia coli.  
 CC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 CC Escherichia.  
 CC NCBI\_TaxID=562;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA STRAIN=PB176;  
 RC MEDLINE=90158116; PubMed=2576094;  
 RA Jatajakumari M.B., Thomas C.J., Halter R., Manning P.A.;  
 RT "Genes for biosynthesis and assembly of CS3 pili of CFA/II  
 RT enterotoxigenic Escherichia coli: novel regulation of pilus  
 RT production by bypassing an amber codon.";  
 RL Mol. Microbiol. 3:1685-1695(1989).  
 CC -1- FUNCTION: THIS PROTEIN IS ESSENTIAL FOR THE BIOGENESIS OF

CC CC MATRE CS3 PILI.  
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. OUTER MEMBRANE  
 CC (BY SIMILARITY).  
 CC -1- ALTERNATIVE PRODUCTS: FIVE PROTEIN ARE PRODUCED BY ALTERNATIVE  
 CC INITIATION. THE FIFTH PROTEIN REQUIRES THE SUPPRESSION OR  
 CC REATHROUGH OF AN INTERNAL AMBER CODON IN POSITION 734.  
 CC -1- SIMILARITY: BELONGS TO THE FIMBRIAL EXPORT USHER FAMILY.  
 CC -----  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: X16944; CAA34817.1; -  
 DR EMBL: X16944; CAA34816.1; ALT\_INIT.  
 DR InterPro: IPR000015; -  
 DR Pfam: PF00577; Usher; 1  
 DR PROSITE: PS01151; FIMBRIAL\_USHER; 1.  
 KM Outer membrane; Transmembrane; Fimbria; Transport; Signal;  
 KM Alternative initiation.  
 FT SIGNAL 1 ?  
 FT CHAIN ? 937 POTENTIAL.  
 FT CHAIN 181 753 CS3 PILI SYNTHESIS 104 KDA PROTEIN.  
 FT CHAIN 317 753 CS3 PILI SYNTHESIS 63 KDA PROTEIN.  
 FT CHAIN 451 753 CS3 PILI SYNTHESIS 48 KDA PROTEIN.  
 FT CHAIN 572 753 CS3 PILI SYNTHESIS 33 KDA PROTEIN.  
 FT CHAIN 572 753 CS3 PILI SYNTHESIS 20 KDA PROTEIN.  
 SQ SEQUENCE 937 AA; 104150 MW; 6C35056EE96277B5 CRC64;  
 Query Match 82.9%; Score 34; DB 1; Length 937;  
 Best Local Similarity 75.0%; Pred. No. 41;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 Oy 1 GGVLYODG 8  
 Db 553 GGLVYDGG 560  
 RESULT 7  
 E13B\_PEA STANDARD: PRT; 370 AA.  
 ID E13B\_PEA  
 AC Q03467;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 01-OCT-1996 (Rel. 34, Last annotation update)  
 DE GLUCAN ENDO-1,3-BETA-GLUCOSIDASE PRECURSOR (EC 3.2.1.39) ((1->3)-  
 DE BETA-GLUCAN ENDOHYDROLASE) ((1->3)-BETA-GLUCANASE) (BETA-1,3-  
 DE ENDOGLUCANASE).  
 OS Pisum sativum (Garden pea).  
 CC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 CC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;  
 CC Fabales; Fabaceae; Papilionoideae; Pisum.  
 CC NCBI\_TaxID=3888;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA STRAIN=CV. ALCAN; TISSUE=Leaf;  
 RC MEDLINE=94143473; PubMed=8310053;  
 RA Chang M.-M., Cullley D.E., Hadwiger L.A.;  
 RT "Nucleotide sequence of a pea (Pisum sativum L.) beta-1,3-glucanase  
 RT gene.";  
 RL Plant Physiol. 101:1121-1122(1993).  
 RN [2]  
 RP SEQUENCE OF 32-370 FROM N.A.  
 RA STRAIN=CV. ALCAN; TISSUE=Leaf;  
 RC MEDLINE=93081722; PubMed=1450378;  
 RA Chang M.-M., Hadwiger L.A., Horovitz D.;  
 RT "Molecular characterization of a pea beta-1,3-glucanase induced by  
 RT Fusarium solani and chitosan challenge.";  
 RL Plant Mol. Biol. 20:609-618(1992).  
 CC -1- FUNCTION: IMPLICATED IN THE DEFENSE OF PLANTS AGAINST PATHOGENS.

```
CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF 1,3-BETA-D-GLUCOSIDIC LINKAGES
CC IN 1,3-BETA-D-GLUCANS.
CC -1- TISSUE SPECIFICITY: CONSTITUTIVELY EXPRESSED IN SEEDLING ROOTS.
CC -1- INDUCTION: BY FUNGAL ELICTORS.
CC -1- SIMILARITY: BELONGS TO FAMILY 17 OF GLYCOSYL HYDROLASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; S51479; AAB24398.1; -.
CC EMBL; L02212; AAA33648.1; -.
CC EMBL; S69419; -. NOT_ANNOTATED_CDS.
CC HSSP; P15737; 1GHS.
CC Interpro: IPR000490; -.
CC Pfam; PF00332; Glyco_hydro_17; 1.
CC PROSITE; PS00587; GLYCOSYL_HYDROL_F17; 1.
CC K0 Hydrolyase; Glycosidase; Glycoprotein; signal.
CC SIGNAL 1 32 POTENTIAL.
CC CHAIN 33 347 GLUCAN ENDO-1,3-BETA-GLUCOSIDASE.
CC PROPEP 348 370 REMOVED IN MATURE FORM (BY SIMILARITY).
CC MOD_RES 33 33 PYROGLUTAMINE CARBOXYLIC ACID (PROBABLE).
CC ACT_SITE 272 272 NUCLEOPHILE (BY SIMILARITY).
CC ACT_SITE 328 328 PROTON DONOR (BY SIMILARITY).
CC CARBOHYD 361 361 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC SEQUENCE 370 AA; 41049 MW; 93847A4CF99B6186 CRC64;

Query Match 80.5%; Score 33; DB 1; Length 370;
Best Local Similarity 85.7%; Pred. No. 25;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy 2 GVLVODG 8
Db 233 GVMVODG 239

RESULT 8
A2BP_MOUSE STANDARD; PRT: 396 AA.
AC Q9J4J3; Q9J4J5;
DT 01-OCT-2000 (Rel. 40, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE ATRAXIN 2-BINDING PROTEIN.
GN A2BP1 OR A2BP.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RP NCB1
RN NCB1
RT "Isolation of full-length cDNA clones from mouse brain cDNA library
RT made by oligo-capping method."
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Brain;
RA Osada N., Kusuda J., Yanuma R., Ito A., Hirata M., Sugano S.,
RA Hashimoto K.;
RT "A novel protein with RNA-binding motifs interacts with ataxin-2.";
RL Hum. Mol. Genet. 9:1303-1313(2000).
CC -1- SUBUNIT: BINDS TO THE C-TERMINAL OF ATRAXIN-2.
CC -1- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RRM).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AK001027; BAA91472.1; -.
CC EMBL; AF107203; AAF78291.1; -.
CC MIM; 605104; -.
CC PROSITE; PS50102; RRM; 1.
```

```
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AB041596; BAA95079.1; -.
CC EMBL; AF107204; AAF78292.1; -.
CC PROSITE; PS50102; RRM; 1.
CC PROSITE; PS00303; RRM_RNP_1; 1.
CC K0 RNA-binding.
CC DOMAIN 116 192 RNA-BINDING (RRM).
CC CONFLICT 60 60 G -> D (IN REF. 2).
CC CONFLICT 128 128 R -> RFR (IN REF. 2).
CC CONFLICT 369 369 G -> S (IN REF. 2).
CC SEQUENCE 396 AA; 42590 MW; 9A0C59C5E687F39F CRC64;

Query Match 80.5%; Score 33; DB 1; Length 396;
Best Local Similarity 75.0%; Pred. No. 27;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Oy 1 GVLVODG 8
Db 295 GGVVODG 302

RESULT 9
A2BP_HUMAN STANDARD; PRT: 397 AA.
AC Q9MWB1; Q9NS20;
DT 01-OCT-2000 (Rel. 40, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE ATRAXIN 2-BINDING PROTEIN.
GN A2BP1 OR A2BP.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RP NCB1
RN NCB1
RT "NEBO human cDNA sequencing project."
RT Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 21-397 FROM N.A.
RC MEDLINE=20275619; PubMed=10814712;
RA Shibata H., Huynh D.P., Puist S.-M.;
RT "A novel protein with RNA-binding motifs interacts with ataxin-2.";
RL Hum. Mol. Genet. 9:1303-1313(2000).
CC -1- SUBUNIT: BINDS TO THE C-TERMINAL OF ATRAXIN-2.
CC -1- TISSUE SPECIFICITY: PREDOMINANTLY EXPRESSED IN MUSCLE AND BRAIN.
CC -1- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RRM).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AK001027; BAA91472.1; -.
CC EMBL; AF107203; AAF78291.1; -.
CC MIM; 605104; -.
CC PROSITE; PS50102; RRM; 1.
```

DR PROSITE; PS00030; RRM\_INP-1; 1.  
 KW RNA-binding. 117 193 RNA-BINDING (RRM).  
 FT DOMAIN 92 A -> T (IN REF. 2).  
 FT CONFLICT 92  
 SQ SEQUENCE 397 AA; 42754 MW; E3E9060B68C79880 CRC64;

Query Match 80.5%; Score 33; DB 1; Length 397;  
 Best Local Similarity 75.0%; Pred. No. 27;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGVYVODG 8  
 DB 296 GGVYVODG 303

RESULT 10  
 ID CHIA\_ECOLI STANDARD; PRT; 897 AA.  
 AC P13656;  
 DT 01-JAN-1990 (Rel. 13, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE PROBABLE BIFUNCTIONAL CHITINASE/LYSOZYME PRECURSOR [INCLUDES:  
 CHITINASE (EC 3.2.1.14); LYSOZYME (EC 3.2.1.17)].  
 GN CHIA.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Escherichia.  
 OX NCBI\_TaxID=562;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-K12 / MG1655;  
 RX MEDLINE=97426517; PubMed=9278503;  
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
 RA Mau B., Shao Y.;  
 RT "The complete genome sequence of Escherichia coli K-12";  
 RT Science 277:1143-1474(1997).  
 RN [2]  
 RP SEQUENCE OF 874-897 FROM N.A.  
 RC STRAIN-K12;  
 RX MEDLINE=89291745; PubMed=2661540;  
 RA Andrews S.C., Harrison P.M., Guest J.R.;  
 RT "Cloning, sequencing, and mapping of the bacterioferritin gene (bfr)  
 RT of Escherichia coli K-12";  
 RL J. Bacteriol. 171:3940-3947(1989).  
 RN [3]  
 RP CHITIN-BINDING PROPERTIES.  
 RX MEDLINE=99350427; PubMed=10419961;  
 RA Simpson H.D., Barras F.;  
 RT "Functional analysis of the carbohydrate-binding domains of Erwina  
 RT chrysanthemi Cels (Endoglucanase Z) and an Escherichia coli putative  
 RT chitinase";  
 RL J. Bacteriol. 181:4611-4616(1999).  
 RN [4]  
 RP CHARACTERIZATION.  
 RC STRAIN-K12 / MG1655;  
 RX MEDLINE=20223635; PubMed=10760150;  
 RA Francetic O., Badout C., Rimsley S., Pugsley A.P.;  
 RT "The ChIA (YheB) protein of Escherichia coli K-12 is an endochitinase  
 RT whose gene is negatively controlled by the nucleoid-structuring  
 RT protein H-NS";  
 RL Mol. Microbiol. 35:1506-1517(2000).  
 CC -1- FUNCTION: BIFUNCTIONAL ENZYME WITH LYSOZYME/CHITINASE ACTIVITY.  
 CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF THE 1,4-BETA-LINKAGE OF  
 CC N-ACETYL-D-GLUCOSAMINE POLYMERS OF CHITIN  
 CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF THE 1,4-BETA-LINKAGES BETWEEN  
 CC N-ACETYL-D-GLUCOSAMINE AND N-ACETYLMURAMIC ACID IN PEPTIDOGLYCAN  
 CC HETEROPOLYMERS OF THE PROKARYOTES PERIPLASMIC.  
 CC -1- SUBCELLULAR LOCATION: PERIPLASMIC.  
 CC -1- MISCELLANEOUS: IN WILD-TYPE E. COLI K12, CHIA EXPRESSION IS

CC REPRODUCED. THE DNA-BINDING PROTEIN H-NS PARTICIPATES IN THE  
 CC SILENCING OF CHIA.  
 CC -1- SIMILARITY: IN THE C-TERMINAL SECTION, BELONGS TO CHITINASE CLASS  
 CC II (FAMILY 18 OF GLYCOSYL HYDROLASES).  
 CC -1- SIMILARITY: CONTAINS FIVE COPIES OF A CHITIN-BINDING DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation-  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL; U18997; AAA58135.1; -;  
 DR EMBL; AE000410; AAC76363.1; -;  
 DR EMBL; M27176; AAC13985.1; -;  
 DR PIR; PV0012; PV0012.  
 DR Ecogene; EG11237; CHIA.  
 DR InterPro; IPR001579; -;  
 DR PROSITE; PS01095; CHITINASE\_18; FALSE NEG.  
 KW Hydrolase; Glycosidase; Chitin degradation; Chitin-binding; Signal;  
 KW Repeat; Periplasmic; Multifunctional enzyme.  
 FT SIGNAL 1 24  
 FT CHAIN 25 897 PROBABLE BIFUNCTIONAL CHITINASE/LYSOZYME.  
 FT DOMAIN 93 101 POLY-SER.  
 FT DOMAIN 25 91 CHITIN-BINDING 1.  
 FT DOMAIN 128 194 CHITIN-BINDING 2.  
 FT DOMAIN 229 295 CHITIN-BINDING 3.  
 FT DOMAIN 337 403 CHITIN-BINDING 4.  
 FT DOMAIN 459 529 CHITIN-BINDING 5.  
 FT DOMAIN 647 897 CATALYTIC (BY SIMILARITY).  
 FT ACT\_SITE 700 700 PROTON DONOR (BY SIMILARITY).  
 FT DISULFID 628 673 BY SIMILARITY.  
 FT CONFLICT 874 874 F -> I (IN REF. 2).  
 SQ SEQUENCE 897 AA; 97057 MW; 96BD145BA1F954F3 CRC64;

Query Match 80.5%; Score 33; DB 1; Length 897;  
 Best Local Similarity 75.0%; Pred. No. 62;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGVYVODG 8  
 DB 407 GGVYVODG 414

RESULT 11  
 ID GUAN\_PIG STANDARD; PRT; 109 AA.  
 AC P79897;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 15-JUL-1998 (Rel. 36, Last annotation update)  
 DE GUANYLIN PRECURSOR (GUANYLATE CYCLASE ACTIVATOR 2A).  
 GN GUCA2A OR GUCA2.  
 OS Sus scrofa (Pig).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.  
 OX NCBI\_TaxID=9823;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Jejunum;  
 RA Hill O., Zucht H.D., Luttermann D., Maegerl H.J., Forssmann W.-G.;  
 RT Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: ENDOGENOUS ACTIVATOR OF INTESTINAL GUANYLATE CYCLASE.  
 CC IT STIMULATES THIS ENZYME THROUGH THE SAME RECEPTOR BINDING REGION  
 CC AS THE HEAT-STABLE ENTEROTOXINS.  
 CC -1- SIMILARITY: BELONGS TO THE GUANYLIN FAMILY. ALSO SIMILAR TO HEAT-  
 CC STABLE ENTEROTOXINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -



CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.1sb-sib.ch/announce/>  
 CC or send an email to [license@1sb-sib.ch](mailto:license@1sb-sib.ch)).

DR EMBL: Z73607; CAA97974.1; -  
 DR HSPB; Q02747; 1GNB.  
 DR InterPro: IPR000879; -  
 DR Pfam: PF02058; Guanylin; 1.  
 DR PRINTS: PRO0774; GUANYLIN.  
 KW Signal.  
 FT SIGNAL. 1 21 POTENTIAL.  
 FT PEPTIDE 95 109 GUANYLIN.  
 FT DISULFID 98 106 BY SIMILARITY.  
 FT DISULFID 101 109 BY SIMILARITY.  
 SQ SEQUENCE 109 AA; 11938 MW; 6368BD1EB0B07A74 CRC64;

Query Match 78.0% Score 32; DB 1; Length 109;  
 Best Local Similarity 75.0% Pred. No. 11;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 GGVLVQDG 8  
 ||| |||  
 Db 20 GGVTVKDG 27

RESULT 12  
 RFA2\_CRIFA STANDARD; PRT: 258 AA.  
 ID RFA2\_CRIFA  
 AC Q23697;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE REPLICATION PROTEIN A 28 KDA SUBUNIT (RP-A) (REPLICATION  
 DE FACTOR-A PROTEIN 2).  
 GN RPA2.  
 OS Crithidia fasciculata.  
 OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Crithidia.  
 OX NCBI\_TaxID=5556;  
 RN [1]  
 RC SEQUENCE FROM N.A.  
 RP STRAIN=CECL1;  
 RX MEDLINE=94239461; PubMed=8183313;  
 RA Brown G.W., Hines J.C., Fisher P., Ray D.S.;  
 RT "Isolation of the genes encoding the 51-kilodalton and 28-kilodalton  
 RT subunits of Crithidia fasciculata replication protein A.";  
 RL Mol. Biochem. Parasitol. 63:135-142(1994).  
 CC -1- FUNCTION: ABSOLUTELY REQUIRED FOR SIMIAN VIRUS 40 DNA REPLICATION  
 CC IN VITRO. IT PARTICIPATES IN A VERY EARLY STEP IN INITIATION. RP-A  
 CC IS A SINGLE-STRANDED DNA-BINDING PROTEIN.  
 CC -1- SUBUNIT: HETEROIMER OF 51, 28, AND 14 KDA CHAINS.  
 CC -1- SIMILARITY: TO OTHER SPECIES RFA2/RPA2.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.1sb-sib.ch/announce/>  
 CC or send an email to [license@1sb-sib.ch](mailto:license@1sb-sib.ch)).

DR EMBL: Z23164; CAA80683.1; -  
 DR DNA replication; Nuclear protein.  
 SQ SEQUENCE 258 AA; 27509 MW; 739EF6D581C8B8E6 CRC64;

Query Match 78.0% Score 32; DB 1; Length 258;  
 Best Local Similarity 75.0% Pred. No. 27;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 GGVLVQDG 8  
 ||| |||  
 Db 46 GGVTVVQDG 53

RESULT 13  
 VG02\_BPM15 STANDARD; PRT: 259 AA.  
 ID VG02\_BPM15  
 AC Q05230;  
 DT 01-FEB-1994 (Rel. 28, Created)  
 DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DT 01-FEB-1994 (Rel. 28, Last annotation update)  
 DE GENE 2 PROTEIN (GP2).  
 GN 2.  
 OS Mycobacteriophage L5.  
 OS Viruses.  
 OX NCBI\_TaxID=12376;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=93211282; PubMed=8459766;  
 RA Hatfull G.F., Sarkis G.J.;  
 RT "DNA sequence, structure and gene expression of mycobacteriophage L5:  
 RT a phage system for mycobacterial genetics.";  
 RL Mol. Microbiol. 7:395-405(1993).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.1sb-sib.ch/announce/>  
 CC or send an email to [license@1sb-sib.ch](mailto:license@1sb-sib.ch)).

DR EMBL: Z18946; CAA79381.1; -  
 DR PIR: S30950; S30950.  
 SQ SEQUENCE 259 AA; 28678 MW; 4B927C38272E102C CRC64;

Query Match 78.0% Score 32; DB 1; Length 259;  
 Best Local Similarity 85.7% Pred. No. 27;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GGVLVQDG 7  
 ||| |||  
 Db 225 GGVLVQDG 231

RESULT 14  
 YKA2\_CAEEL STANDARD; PRT: 273 AA.  
 ID YKA2\_CAEEL  
 AC P34254;  
 DT 01-FEB-1994 (Rel. 28, Created)  
 DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DT 01-FEB-1994 (Rel. 28, Last annotation update)  
 DE HYPOTHETICAL 30.9 KDA PROTEIN B0303.2 IN CHROMOSOME III.  
 GN B0303.2.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;  
 OC Rhabditidae; Pelodierinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BRISTOL N2;  
 RX MEDLINE=92168156; PubMed=1538779;  
 RA Sulston J., Du Z., Thomas K., Wilson R., Hillier L., Staden R.,  
 RA Halloran N., Green P., Thierry-Mieg J., Olf L., Dear S., Coulson A.,  
 RA Craxton M., Durbin R.K., Berks M., Metzstein M., Hawkins T.,  
 RA Ainscough R., Waterston R.;  
 RT "The C. elegans genome sequencing project: a beginning.";  
 RL Nature 356:37-41(1992).  
 CC -1- SIMILARITY: STRONG, TO MAMMALIAN PHENYLETHANOLAMINE  
 CC N-METHYLTTRANSFERASE (EC 2.1.1.28).  
 CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; M77697; AAA27896.1; ALT\_SEQ.  
 DR PIR; S27784.  
 DR WormPep; B0303.2; CE00336.  
 DR InterPro; IPR000940;  
 DR Pfam; PF01234; NMMT\_PMTT\_TEMT; 1.  
 DR Hypothetical protein; Transferase; Methyltransferase.  
 KM SEQUENCE 273 AA; 30060 MW; 11A793EC49C270C0 CRC64;

Query Match 7/11.0%; Score 32; DB 1; Length 273;  
 Best Local Similarity 67.5%; Pred. No. 29;  
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 GGVLYODG 8  
 ||:|:|  
 Db 209 GGILIGG 216

RESULT 15  
 ID METE\_METTM STANDARD; PRT; 309 AA.  
 AC P55239;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-2000 (Rel. 40, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE PROBABLE METHYLCOBALAMIN:HOMOCYSTEINE METHYLTRANSFERASE (EC 2.1.1.-)  
 DE (METHIONINE SYNTHASE).  
 GN METE.  
 OS Methanobacterium thermocautotrophicum (strain Marburg / DSM 2133).  
 OC Archaea; Euryarchaeota; Methanobacteriales; Methanobacteriaceae;  
 OC Methanothermobacter.  
 OX NCBI\_TaxID=79929;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE-96184910; PubMed-8617278;  
 RA Vaupel M., Dietz H., Linder D., Thauer R.K.;  
 RT "Primary structure of cyclohydrolase (Mch) from Methanobacterium  
 RT thermocautotrophicum (strain Marburg) and functional expression of the  
 RT mch gene in Escherichia coli.";  
 RL Eur. J. Biochem. 236:294-300(1996).  
 RN [2]  
 RP SEQUENCE OF 3-24, AND CHARACTERIZATION.  
 RA MEDLINE-99398345; PubMed-10469143;  
 RA Schroeder I., Thauer R.K.;  
 RT "Methylcobalamín:homocysteine methyltransferase from Methanobacterium  
 RT thermocautotrophicum. Identification as the mch gene product.";  
 RL Eur. J. Biochem. 263:789-796(1999).  
 CC -1- FUNCTION: CATALYZES THE TRANSFER OF A METHYL GROUP FROM  
 CC METHYLCOBALAMIN AND METHYLCOBINAMIDE TO HOMOCYSTEINE RESULTING IN  
 CC METHIONINE FORMATION.  
 CC -1- COFACTOR: ZINC. BINDS ONE MOLE PER SUBUNIT (BY SIMILARITY).  
 CC -1- PATHWAY: TERMINAL STEP IN THE DE NOVO BIOSYNTHESIS OF METHIONINE.  
 CC -1- SUBUNIT: MONOMER.  
 CC -1- SIMILARITY: BELONGS TO THE VITAMIN-B12 INDEPENDENT METHIONINE  
 CC SYNTHASE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR InterPro: IPR002629;  
 DR Pfam: PF01717; Methionine\_Synt; 1.  
 KM Transferase; Methyltransferase; Methionine biosynthesis; Zinc.  
 FT METAL 201 201 ZINC (BY SIMILARITY).  
 FT METAL 203 203 ZINC (BY SIMILARITY).  
 FT METAL 285 285 ZINC (BY SIMILARITY).  
 FT METAL 285 285 ZINC (BY SIMILARITY).  
 SO SEQUENCE 309 AA; 33532 MW; 522996A39DF1E31A CRC64;

Query Match 78.0%; Score 32; DB 1; Length 309;  
 Best Local Similarity 62.5%; Pred. No. 33;  
 Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 1 GGVLYODG 8  
 ||:|:|  
 Db 70 GGMYVADG 77

Search completed: June 13, 2001, 14:21:47  
 Job time: 807 sec



*This Page Blank (uspto)*

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 13, 2001, 14:20:34 ; Search time 225.85 seconds  
(without alignments)  
4.152 Million cell updates/sec

Title: PCT-US01-05825A-16

Perfect score: 41

Sequence: 1 GGVLVQDG 8

Scoring table:

BLOSUM62

Gapped 10.0 , Gapext 0.5

Searched: 374700 seqs, 117207915 residues

Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: SP archaea:\*  
2: SP bacteria:\*  
3: SP fungi:\*  
4: SP human:\*  
5: SP\_invertebrate:\*  
6: SP\_mammal:\*  
7: SP\_mhc:\*  
8: SP\_organelle:\*  
9: SP\_phage:\*  
10: SP\_plant:\*  
11: SP\_rodent:\*  
12: SP\_unclassified:\*  
13: SP\_vertebrate:\*  
14: SP\_virus:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	38	92.7	1491	10	Q9XZ6 oryza sativ
2	37	90.2	1542	10	Q9ZS84 lycopersico
3	35	85.4	537	2	005459 mycobacteri
4	35	85.4	606	2	Q9RX21 deinococcus
5	35	85.4	1510	5	061802 caenorhabdi
6	34	82.9	250	5	Q9NN02 leishmania
7	34	82.9	303	2	P97157 escherichia
8	34	82.9	476	10	004927 allium tude
9	34	82.9	538	2	068180 uncultured
10	34	82.9	1241	10	Q9LHK4 arabidopsis
11	34	82.9	1587	10	Q9SLQ0 mycobacteri
12	33	80.5	157	2	P96324 oryza sativ
13	33	80.5	213	2	Q9RW35 deinococcus
14	33	80.5	218	2	P94935 micrococcus
15	33	80.5	288	10	Q9SN11 oryza sativ
16	33	80.5	355	2	046210 chlamydomon
17	33	80.5	368	10	Q40314 medicago sa
18	33	80.5	377	4	Q9NS20 homo sapien
19	33	80.5	378	11	Q9J43 mus musculu

20	33	80.5	396	11	Q9JRB5 mus musculu
21	33	80.5	397	4	Q9NMB1 homo sapien
22	33	80.5	418	2	P77552 escherichia
23	33	80.5	522	3	P74199 metathizium
24	33	80.5	569	1	074030 methanobact
25	33	80.5	765	3	014425 aspergillus
26	33	80.5	1240	10	Q9LHDL arabidopsis
27	33	80.5	1335	2	Q9LA58 escherichia
28	33	80.5	1335	2	Q9LA54 escherichia
29	32	78.0	122	10	Q9SBY5 oryza sativ
30	32	78.0	134	2	066548 aquifex aeo
31	32	78.0	147	10	Q9ZTK2 vitis vinif
32	32	78.0	147	10	024551 vitis vinif
33	32	78.0	242	1	029577 archaeoglob
34	32	78.0	322	10	Q9SGL1 musa acumin
35	32	78.0	340	10	022317 musa acumin
36	32	78.0	345	10	Q9M3U4 vitis vinif
37	32	78.0	529	2	Q9L9H7 salmonella
38	32	78.0	530	2	Q9KVR0 oryza sativ
39	32	78.0	620	10	Q9SR3 caenorhabdi
40	32	78.0	1034	5	Q20398 caenorhabdi
41	32	78.0	1302	5	Q9VRW3 dirosophila
42	31	75.6	40	2	09L171 streptomyce
43	31	75.6	103	10	022062 dioscorea t
44	31	75.6	119	10	024620 dioscorea t
45	31	75.6	119	10	024622 dioscorea t

## ALIGNMENTS

RESULT 1  
Q9XZ6 PRELIMINARY; PRT; 1491 AA.  
AC Q9XZ6;  
DT 01-NOV-1999 (TREMUREL.12, Created)  
DT 01-NOV-1999 (TREMUREL.12, Last sequence update)  
DT 01-JUN-2000 (TREMUREL.14, Last annotation update)  
DE HYPOTHEICAL 170.8 KDA PROTEIN.  
OS Oryza sativa (Rice).  
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
OC Magnoliophyta; Liliopsida; Poales; Poaceae; Oryza.  
OX NCBI\_TaxID=4530;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CV. NIPPONBARE;  
RA Sasaki T., Matsumoto T., Yamamoto K.;  
RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 8, PAC  
RT clone:PO026F07." to the EMBL/GenBank/DBJ databases.  
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AP000364; BAA81778.1; -  
DR INTERPRO: IPR000477; -  
DR INTERPRO: IPR001878; -  
DR INTERPRO: IPR001969; -  
DR PRAM: PR00078; TVC; 1.  
DR PRAM: PR00098; ZF-CCHC; 1.  
DR PROSITE: PS00141; ASP\_PROTEASE; UNKNOWN\_1.  
KW Hypothetical protein.  
SQ SEQUENCE 1491 AA; 170789 MW; 7DABC98047FC409 CRC64;

Query Match 92.7%; Score 38; DB 10; Length 1491;  
Best local similarity 87.5%; Pred. No. 61;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGVLVQDG 8  
DB 1035 GGVLVQDG 1042  
RESULT 2  
Q9ZS84 PRELIMINARY; PRT; 1542 AA.

AC 092584;  
 DT 01-MAY-1999 (TREMBlrel. 10, Created)  
 DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)  
 DT 01-JUN-2000 (TREMBlrel. 15, Last annotation update)  
 DE POLYPROTEIN.  
 OS Lycopersicon esculentum (Tomato).  
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 OC Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I;  
 OC Solanales; Solanaceae; Solanum.  
 NCBI\_TaxID=4081;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CV. MONEYMAKER;  
 RA Parniske M., Jones J.D.G.;  
 RT "Recombination between diverged clusters of the tomato Cf-9 plant  
 disease resistance gene family";  
 RL Proc. Natl. Acad. Sci. U.S.A. 0:0-0(1999).  
 [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CV. MONEYMAKER;  
 RA MEDLINE-99125234; PubMed-9926411;  
 RA Parniske M., Wulff B.B., Bonnem G., Thomas C.M., Jones D.A.,  
 RA Jones J.D.;  
 RT "Homologues of the Cf-5 disease resistance gene (Hc9s) are present at  
 multiple loci on the short arm of tomato chromosome 1";  
 RL Mol. Plant Microbe Interact. 12:93-102(1999).  
 DE EMBL; AF119040; AAD13304.1; -;  
 DR INTERPRO; IPR000477; -;  
 DR INTERPRO; IPR000953; -;  
 DR INTERPRO; IPR001584; -;  
 DR INTERPRO; IPR001669; -;  
 DR PFM; PF00078; Tve; 1.  
 DR PFM; PF00665; Tve; 1.  
 DR PROSITE; PS00141; ASP\_PROTEASE; UNKNOWN\_1.  
 DR PROSITE; PS0013; CHROMO\_2; 1.  
 KW Polyprotein.  
 SQ SEQUENCE 1542 AA; 175610 MW; OCC1A1A1A1455670 CRC64;

Query Match 90.2%; Score 37; DB 10; Length 1542;  
 Best Local Similarity 87.5%; Pred. No. 1e+02;  
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGVLYODG 8  
 |||||:|  
 DB 920 GGVLYODG 927

RESULT 3  
 ID 005459 PRELIMINARY; PRT; 537 AA.  
 AC 005459;  
 DT 01-JUL-1997 (TREMBlrel. 04, Created)  
 DT 01-JUL-1997 (TREMBlrel. 04, Last sequence update)  
 DT 01-JUN-2000 (TREMBlrel. 14, Last annotation update)  
 DE HYPOTHEICAL 57.7 KDA PROTEIN.  
 GN RV3885C OR MFCY15P10.27.  
 OS Mycobacterium tuberculosis.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.  
 NCBI\_TaxID=1773;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-H37RV;  
 RX MEDLINE-98295987; PubMed-9634230;  
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
 RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,  
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,  
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,  
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R., Sulston J.E.,  
 RA Taylor K., Whitehead S., Barrett B.G.,

RT "Deciphering the biology of Mycobacterium tuberculosis from the  
 complete genome sequence.";  
 RL Nature 393:537-544(1998).  
 DE EMBL; Z94121; CAB08089.1; -;  
 DR TUBERCULIST; RV3885C; -;  
 KW Hypothetical protein.  
 SQ SEQUENCE 537 AA; 57656 MW; 9593AD9DA6DD03E5 CRC64;

Query Match 85.4%; Score 35; DB 2; Length 537;  
 Best Local Similarity 87.5%; Pred. No. 74;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGVLYODG 8  
 |||||:  
 DB 87 GGVLYODG 94

RESULT 4  
 ID 09RXZ1 PRELIMINARY; PRT; 606 AA.  
 AC 09RXZ1;  
 DT 01-MAY-2000 (TREMBlrel. 13, Created)  
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
 DT 01-JUN-2000 (TREMBlrel. 14, Last annotation update)  
 DE ABC TRANSPORTER, ATP-BINDING PROTEIN, MSBA FAMILY.  
 GN DR0163.  
 OS Deinococcus radiodurans.  
 OC Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.  
 NCBI\_TaxID=1299;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-R1;  
 RX MEDLINE-20036896; PubMed-10567266;  
 RA White O., Eissen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,  
 RA Dodson R.V., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,  
 RA Morfitt K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,  
 RA Vamathevan J.J., Lam P., McDonald L., Urdaneta T., Zalewski C.,  
 RA Makarewicz K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,  
 RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,  
 RA Fraser J.M.;  
 RT "Genome Sequence of the Radioresistant Bacterium Deinococcus  
 radiodurans R1.";  
 RL Science 286:1571-1577(1999).  
 DR EMBL; AE001878; AAF09749.1; -;  
 DR TIGR; DR0163; -;  
 DR INTERPRO; IPR001066; -;  
 DR INTERPRO; IPR001140; -;  
 DR INTERPRO; IPR001617; -;  
 DR PFM; PF00005; ABC\_tran; 1.  
 DR PFM; PF00664; ABC\_membrane; 1.  
 DR PROSITE; PS00211; ABC\_TRANSPORTER; 1.  
 DR PROSITE; PS00216; SUGAR\_TRANSPORT\_1; UNKNOWN\_1.  
 KW ATP-binding.  
 SQ SEQUENCE 606 AA; 65091 MW; 4AC133D5FA23D0AC CRC64;

Query Match 85.4%; Score 35; DB 2; Length 606;  
 Best Local Similarity 75.0%; Pred. No. 85;  
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGVLYODG 8  
 |||||:  
 DB 561 GGVLYODG 568

RESULT 5  
 ID 061802 PRELIMINARY; PRT; 1510 AA.  
 AC 061802;  
 DT 01-AUG-1998 (TREMBlrel. 07, Created)  
 DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)  
 DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)

DE H1LE01.3 PROTEIN.  
CN H1LE01.3.  
OS *Caenorhabditis elegans*.  
CC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidae;  
OC Rhabditidae; Pelodermineae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RX MEDLINE=94150718; PubMed=7906398;  
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,  
RA Bonfield J., Burton J., Connell M., Copey T., Cooper J., Coulson A.,  
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,  
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnson L.,  
RA Jones M., Kershaw J., Kirsten J., Lalister N., Latelle P.,  
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,  
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R.,  
RA Smidson N., Smith A., Sonhammer E., Staden R., Sulston J.,  
RA Thierfey-Mieg J., Thomas K., Vaundt M., Vaughan K., Waterston R.,  
RA Watson A., Weinstock L., Wilkison-Sproat J., Wohldman P.;  
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of *C.*  
RL *elegans*."; Nature 368:32-38(1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RA Geisel C., Mameley P.;  
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RA Waterston R.;  
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.  
DN EMBL AF067222; AAC17017.1; --  
SQ SEQUENCE 1510 AA; 166889 MW; C1C08ACBC6D6FF11 CRC64;

Query Match	Best Local Similarity	Score 35; DB 5; Length 1510; 85.4%; 75.0%; Pred. No. 2.5e+02;
Matches 6;	Conservative 1;	Mismatches 1; Indels 0; Gaps 0;
Oy	1 GGVLVQDG 8	11:1111
Db	1495 GGILVYDGG 1502	
RESULT 6		
O9NN02		
ID O9NN02	PRELIMINARY;	PRT, 250 AA.
AC O9NN02;		
DT 01-OCT-2000 (TREMBLrel, 15, Created)		
DT 01-OCT-2000 (TREMBLrel, 15, Last sequence update)		
DT 01-OCT-2000 (TREMBLrel, 15, Last annotation update)		
DE PROBABLE REPLICATING PROTEIN A 28 KDA SUBUNIT (FRAGMENT).		
LN 15,463.		
OS Leishmania major.		
OC Eukaryota; Elenchozoa; Kinetoplastida; Trypanosomatidae; Leishmania.		
OX NCBI_TaxID=5664;		
[1]		
RN SEQUENCE FROM N.A.		
RP STRAIN=FRIEDLIN;		
RC Murphy L., Quail M., Harris D., Rajandream M., Ivens A., Barrell B.,		
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.		
DR EMBL: AL160371; CAC00408.1; .		
FT NON_TER	1	
FT NON_TER	250	
SEQUENCE	250 AA; DB5688D2625DDBF3 CAC64;	
20		

Query Match	82.9%	Score 34;	DB 5;	Length 250;
Best Local Similarity	87.5%	Pred. NO. 48;		
Matches 7;	Conservative 0;	Mismatches 1;	Indels 0;	Gaps 0;

```

OY      1 GGVLVQDG 8
        |||||
        |||||
        |||||
Db      42 GGVLVVDG 49

RESULT 7
ID      P97157      PRELIMINARY;      PRT;      303 AA.
P97157
AC      P97157;
DT      01-MAY-1997 (TREMBLrel. 03, Created)
DT      01-MAY-1997 (TREMBLrel. 03, Last sequence update)
DT      01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE      33 KDA PROTEIN.
OS      Escherichia coli.
OC      Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC      Escherichia.
OX      NCBI_TaxID=562;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=PB176;
RX      MEDLINE=90158116; PubMed=2576094;
RA      Tajatunumari M.B., Thomas C.J., Halter R., Manning P.A.;
RT      "Genes for biosynthesis and assembly of CS3 pili of CFA/II
RT      enterotoxigenic Escherichia coli: novel regulation of pilus production
RT      by bypassing an amber codon.";
RL      Mol. Microbiol. 3:1685-1695(1989).
DR      EMBL: X16944: CAA34818.1; -.
DR      INTERPRO: IPR000015; -.
PFAM: PF00577; Ushef; 1.
SO      SEQUENCE      303 AA;      32839 MW;      A8C62A5A60164542 CRC64;

```

```

Query Match Similarity      82.9%; Score 34; DB 2; Length 303;
Best Local Similarity      75.0%; Pred. No. 60;
Matches      6; Conservative      1; Indels      0; Gaps      0;

OY      1 GGVLYVDG 8
        ||| |||
Db       103 GGILVYDG 110

RESULT      8
O04927      PRELIMINARY;      PRT;      476 AA.
AC      O04927;
DT      01-JUL-1997 (TREMBLrel. 04, Created)
DT      01-JUL-1997 (TREMBLrel. 04, Last sequence update)
DT      01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE      ALLINASE (EC 4.4.1.4) (ALLIN LYASE) (CYSTEINE SULPHOXIDE LYASE).
OS      Allium tuberosum.
OC      Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC      Magnoliophyta; Liliopsida; Asparagales; Alliaceae; Allium.
OX      NCBI_TaxID=4683;
RN      [1]
RP      SEQUENCE FROM N.A.
RA      Manabe T., Saito K.;
RL      Submitted (May-1997) to the EMBL/GenBank/DDBJ databases.
RC      -1- CATALYTIC ACTIVITY: AN S-ALKYL-L-CYSTEINE S-OXIDE -> AN ALKYL
CC      SULFENATE + 2 AMINOCRYLATE.
CC      -1- COFACTOR: PYRIDOXAL-PHOSPHATE.
CC      EMBL: AB004269; BAA20358.1; -
DR      MENDEL: 16931; ALITu:1125;16931.
DR      INTERPRO: IPR000561; -
DR      PROSITE: PS00022; EGF_1; UNKNOWN_1.
DR      PRODOM: PD014435; - 1.
KW      lyase.
SQ      SEQUENCE      476 AA;      54084 MW;      074B696F559B6492 CRC64;

```

Query Match	82.9%	Score 34;	DB 10;	Length 476;
Best Local Similarity	62.5%;	Pred. No. 1e+02;		
Matches	5;	Conservative 2;	Mismatches 1;	Indels 0;
				Gaps 0;

```
QY      1 GGVLVQDG 8
         ||::|||
Db      413 GGITVQDG 420
```

	RESULT	9			
068180	ID	068180	PRELIMINARY;	PRT;	528 AA.
AC	068180;				
DT	01-AUG-1998 (TREMBLE)	.07, Created)			
DT	01-AUG-1998 (TREMBLE)	.07, last sequence update)			
DT	01-NOV-1998 (TREMBLE)	.08, last annotation update)			
DE	NITROUS OXIDE REDUCTASE (FRAGMENT).				
GN	nosZ.				
OS	uncultured eubacterium S321195C.				
OC	Bacteria; Proteobacteria; environmental samples.				
OX	NCBI_TaxID=65676;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=98258028; PubMed=9595664;				
RA	Scala D.J., Kerkhof L.T.;				
RT	"Nitrous oxide reductase (nosZ) gene-specific PCR primers for detection of denitrifiers and three nosZ genes from marine sediments."				
RL	FEMS Microbiol. Lett. :62:61-68(1988).				
FM	EMBL; AF016057; AAC38341.1;	-.			
FT	NON_TER	1			
FE	NON_TER	528			
SQ	SEQUENCE	528 AA; 58339 MW; CGBA2906B6CEC6BD CRC64;			

	RESULT	10
09LHK4		
ID	09LHK4	PRELIMINARY; PRT; 1241 AA.
AC	09LHK4;	
DT	01-OCT-2000 (TREMBLrel. 15, Created)	
DT	01-OCT-2000 (TREMBLrel. 15, last sequence update)	
DT	01-OCT-2000 (TREMBLrel. 15, last annotation update)	
DE	P-GALACTOPROTEIN, MULTITRANS RESISTANCE RELATED, ABC TRANSPORTER-LIKE PROPEPIN.	
OS	Arabidopsis thaliana (house-ear cress).	
OC	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;	
OC	Magnoliophyta; eudicotyledons: core eudicots; Rosidae; eurosids II;	
OC	Brassicales; Brassicaceae; Arabidopsids.	
OX	NCHI_TaxID=3702;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=COLUMBIA;	
RA	Nakamura Y.;	
RT	*Structural Analysis of Arabidopsis thaliana Chromosome 3. III.*;	
DR	Submitted (May-2000) to the EMBL/Genbank/DBJ databases.	
EMBL:	AP002043; BAB02613.1;	
SQ	SEQUENCE 1241 AA; 137111 MW; DD3367D1E7F07D28 CRC64;	

RESULT	11		
OS	10		
OSL00		PRELIMINARY;	PRT; 1587 AA.
AC	09SL00;		
DT	01-MAY-2000 (Tremblrel, 13, Created)		
DT	01-MAY-2000 (Tremblrel, 13, last sequence update)		
DT	01-OCT-2000 (Tremblrel, 15, last annotation update)		
DE	GAG-POL. POLYPROTEIN (FRAGMENT).		
GN	GAG-POL.		
OS	Oryza sativa (Rice).		
OC	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;		
OC	Magnoliophyta; Liliopsida; Poales; Poaceae; Oryza.		
OX	NCBI_TaxID=4530;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=IR36; TRANSPOSON=GYPSY-TYPE RETROTRANSPOSON RIR67;		
RA	Kunekawa N., Omitido N., Fukui K., Ohtsubo E., Ohtsubo H.;		
RT	"Novel gypsy-type retrotransposon RIR67: insertion sequence into the		
RT	tandem repeat sequence Trsd which localized in pericentromeric		
RT	heterochromatin region in rice.";		
RL	Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.		
DR	EMBL: AB033235. BAAB89466.1; -.		
DR	INTERPRO: IPR000477; -.		
DR	INTERPRO: IPR001584; -.		
DR	INTERPRO: IPR001878; -.		
DR	PFAM: PF00078; rve; 1.		
DR	PFAM: PF00098; zf-CCHC; 1.		
DR	PFAM: PF00665; rve; 1.		
KW	polyprotein.		
FT	NON_TER	1587	1587
CO	SEQUENCE	1587 AA; 180673 MW; 6C2F6BAB2D590969 CRC64;	

RESULT	12	
P96224		
ID	P96224	PRELIMINARY; PRT; 157 AA.
AC	P96224;	
DT	01-MAY-1997 (TEMBLrel. 03, Created)	
DT	01-MAY-1997 (TEMBLrel. 03, Last sequence update)	
DT	01-JUN-2000 (TEMBLrel. 13, Last annotation update)	
DE	HYPOTHEETICAL 16.2 KDA PROTEIN.	
GN	RV3853 OR MTCY01A6.15C.	
OS	Mycobacterium tuberculosis.	
OC	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;	
OC	Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.	
OX	NCBI_TaxID=1773;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=H37Rv.	
RX	MEDLINE=98295987; PubMed=9634220;	
RA	Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,	
RA	Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,	
RA	Davies R.K., Basham D., Brown D., Chillingworth T., Connor R.,	
RA	Baxter C., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,	
RA	Hornsbly T., Jagers K., Krogh A., McLean J., Moule S., Murphy L.,	
RA	Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,	
RA	Taylor K., Whitehead S., Barrrell B.G.;	
RT	"Deciphering the biology of Mycobacterium tuberculosis from the	
RL	complete genome sequence."	
RL	Nature 393:537-544(1998).	
EMBL	Z83864; CAB06213.1; -	
DT	TUBERCULIST; RV3853; -	
DT	Hypothetical protein.	



SO SEQUENCE 157 AA; 16235 MM; 92BD93F60B66E793 CRC64;

Query Match 80.5%; Score 33; DB 2; Length 157;  
Best Local Similarity 87.5%; Pred. No. 44;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GGVLVODG 8  
|||  
Db 61 GGVLVODG 68

RESULT 13

O9RM35 PRELIMINARY; PRT; 213 AA.

AC O9RM35;  
DT 01-MAY-2000 (TREMBLrel. 13, Created)  
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)  
DE NITROGEN REGULATOR, PUTATIVE.  
GN DR0834.

OS Deinococcus radiodurans.  
OC Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.  
OX NCBI\_TaxID=1299;  
RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-R1;

RX MEDLINE=20036896; PubMed=10567266;

RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,

RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,

RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,

RA Vamathevan J.J., Lam P., McDonald L., Utterback T., Zaleski C.,

RA Maratova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,

RA Kechum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,

RA Fraser C.M.;

RT "Genome Sequence of the Radioresistant Bacterium Deinococcus

RT radiodurans R1.";

RL Science 286:1571-1577(1999).

RL EMBL; AE001937; AAF10412.1; -.

DR TIGR; DR0834; -.

DR INTERPRO: IPR000595; -.

DR INTERPRO: IPR001808; -.

DR PFAM: PF00027; CNMP\_binding; 1.

DR PFAM: PF00325; crp; 1.

DR PRINTS: PR00034; HTICRP.

DR PROSITE: PS50042; CNMP\_BINDING\_3; 1.

SO SEQUENCE 213 AA; 23266 MM; A820032A7C252D7 CRC64;

Query Match 80.5%; Score 33; DB 2; Length 213;  
Best Local Similarity 75.0%; Pred. No. 63;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 GGVLVODG 8  
|||  
Db 121 GGVLVODG 128

RESULT 14

P94955 PRELIMINARY; PRT; 218 AA.

AC P94955;

DT 01-MAY-1997 (TREMBLrel. 03, Created)

DT 01-MAY-1997 (TREMBLrel. 03, Last sequence update)

DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)

DE H(+)-TRANSPORTING ATP SYNTHASE (EC 3.6.1.34)

DE (H(+)-TRANSPORTING ATPASE) (MITOCHONDRIAL ATPASE)

DE (CHLOROPLAST ATPASE) (COUPLING FACTORS (F(0), F(1) AND CF(1)))

DE (FRAGMENT).

GN ATP H.

OS Micrococcus luteus (Micrococcus lysodeikticus).

OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;

OC Actinomycetales; Micrococcales; Micrococcaceae; Micrococcus.

OX NCBI\_TaxID=1270;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-ATCC 4698;

RA Gruber G., Dose K., Nawroth T., Diel P., Cattaruzza M., Flach R.;

RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.

CC -1-CATALYTIC ACTIVITY: ATP + H(2)O = ADP + PHOSPHATE.

DR EMBL; X84791; CAA59262.1; -.

DR INTERPRO: IPR000711; -.

DR PFAM: PF00213; OSCP; 1.

DR PRINTS: PR00125; ATPASEDELTA.

KW Hydrolase.

FT NON\_TER

SO SEQUENCE 218 AA; 22568 MM; F085D09081D93381 CRC64;

Query Match 80.5%; Score 33; DB 2; Length 218;  
Best Local Similarity 87.5%; Pred. No. 65;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GGVLVODG 8  
|||  
Db 187 GGVLVODG 194

RESULT 15

O9SN11 PRELIMINARY; PRT; 288 AA.

AC O9SN11;

DT 01-MAY-2000 (TREMBLrel. 13, Created)

DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)

DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)

DE SIMILAR TO MADS-BOX PROTEIN AG14.

OS Oryza sativa (Rice).

OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;

OC Magnoliophyta; Liliopsida; Poales; Poaceae; Oryza.

OX NCBI\_TaxID=4530;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-CV, NIPPONBARE;

RA Sasaki T., Matsumoto T., Yamamoto K.;

RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 6, PAC

RT clone: P0514G12.";

RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.

DR EMBL; AP000616; BAA85428.1; -.

DR HSSP; P11746; IMNM.

DR INTERPRO: IPR002100; -.

DR PFAM: PF00319; SRF-TF; 1.

DR PRINTS: PR00404; MADSDOMAIN.

DR PROSITE: PS50066; MADS\_BOX\_2; 1.

SO SEQUENCE 288 AA; 30675 MM; 88919F5DB8A96623 CRC64;

Query Match 80.5%; Score 33; DB 10; Length 288;  
Best Local Similarity 75.0%; Pred. No. 90;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 GGVLVODG 8  
|||  
Db 214 GGVLVODG 221

Search completed: June 13, 2001, 14:20:35  
Job time: 735 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

## OM protein - protein search, using sw model

Run on: June 13, 2001, 14:16:36 ; Search time 118.55 Seconds

(without alignments)  
1.296 Million cell updates/sec

Title: PCT-US01-05825A-16

Perfect score: 41  
Sequence: 1 GGVLYVDG 8Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 185757 seqs, 19210857 residues

Total number of hits satisfying chosen parameters: 185757

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents, AA:\*

- 1: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep:\*
- 2: /cgn2\_6/ptodata/2/1aa/5B.COMB.pep:\*
- 3: /cgn2\_6/ptodata/2/1aa/6A.COMB.pep:\*
- 4: /cgn2\_6/ptodata/2/1aa/6B.COMB.pep:\*
- 5: /cgn2\_6/ptodata/2/1aa/PCITOS.COMB.pep:\*
- 6: /cgn2\_6/ptodata/2/1aa/Backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	36	87.8	115	2	US-07-903-029-4
2	36	87.8	115	2	US-07-903-029-5
3	34	82.9	20	2	US-08-859-931A-4
4	33	80.5	368	1	US-08-093-327-4
5	32	78.0	401	2	US-08-741-327E-11
6	31	75.6	118	3	US-08-545-809A-125
7	31	75.6	212	1	US-08-716-301-6
8	31	75.6	346	1	US-08-774-169-3
9	30	73.2	115	2	US-07-903-029-6
10	30	73.2	118	3	US-08-545-809A-108
11	30	73.2	119	2	US-08-475-000-16
12	30	73.2	119	2	US-08-483-199-16
13	30	73.2	119	2	US-08-484-508-16
14	30	73.2	123	3	US-08-428-197-38
15	30	73.2	123	5	PCT-US93-10555-38
16	30	73.2	1275	3	US-09-120-513-2
17	30	73.2	1275	3	US-09-450-105-2
18	30	73.2	1408	1	US-08-612-521-2
19	29	70.7	220	4	US-09-004-731-95
20	29	70.7	220	4	US-08-749-699-95
21	29	70.7	226	4	US-09-004-731-8
22	29	70.7	226	4	US-08-749-699-8
23	29	70.7	247	2	US-08-851-974-4
24	29	70.7	247	2	US-09-213-390-4
25	29	70.7	254	4	US-09-004-731-89
26	29	70.7	254	4	US-09-004-731-92
27	29	70.7	254	4	US-08-749-699-89

28	29	70.7	254	4	US-08-749-699-92	Sequence 92, Appl
29	29	70.7	292	3	US-09-109-205-2	Sequence 2, Appl
30	29	70.7	327	4	US-09-004-731-2	Sequence 2, Appl
31	29	70.7	327	4	US-09-004-731-5	Sequence 2, Appl
32	29	70.7	327	4	US-08-749-699-2	Sequence 2, Appl
33	29	70.7	327	4	US-08-749-699-5	Sequence 2, Appl
34	29	70.7	552	3	US-09-120-365-5	Sequence 5, Appl
35	28	68.3	10	3	US-08-159-339A-295	Sequence 295, App
36	28	68.3	15	1	US-08-467-083-32	Sequence 32, Appl
37	28	68.3	15	1	US-08-414-417B-32	Sequence 32, Appl
38	28	68.3	15	2	US-08-486-348A-32	Sequence 32, Appl
39	28	68.3	15	2	US-08-468-545B-32	Sequence 32, Appl
40	28	68.3	15	3	US-08-466-680B-32	Sequence 32, Appl
41	28	68.3	31	3	US-09-045-632-70	Sequence 70, Appl
42	28	68.3	31	3	US-09-045-632-73	Sequence 73, Appl
43	28	68.3	79	3	US-09-100-804-26	Sequence 26, Appl
44	28	68.3	97	1	US-08-421-356-3	Sequence 3, Appl
45	28	68.3	119	2	US-08-318-157B-2	Sequence 2, Appl

## ALIGNMENTS

RESULT 1

US-07-903-029-4

Sequence 4, Application US/07903029

Patent No. 5969097

GENERAL INFORMATION:

APPLICANT: Wiegand, Roger C.

APPLICANT: Currie, Mark C.

TITLE OF INVENTION: Human Guanylin

NUMBER OF SEQUENCES: 6

CORRESPONDENCE ADDRESS:

ADDRESSEE: Dennis A. Bennett, Monsanto Co., A35G

STREET: 800 N. Lindbergh Blvd.

CITY: St. Louis

STATE: Missouri

COUNTRY: USA

ZIP: 63167

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentln Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/903,029

FILING DATE: 19920623

CLASSIFICATION: 530

ATTORNEY/AGENT INFORMATION:

NAME: Bennett, Dennis A.

REGISTRATION NUMBER: 34,547

REFERENCE/DOCKET NUMBER: 07-21(872)A

TELECOMMUNICATION INFORMATION:

TELEPHONE: (314)694-5402

TELEFAX: (314)694-9009

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 115 amino acids

TYPE: AMINO ACID

TOPOLOGY: linear

MOLECULE TYPE: protein

US-07-903-029-4

## Query Match

Best Local Similarity 87.8%;

Matches 7; Conservative 0;

Mismatches 1; Indels 0;

Gaps 0;

Score 36; DB 2; Length 115;

Pred. No. 6.5;

Mismatches 1; Indels 0;

Gaps 0;

RESULT 2  
US-07-903-029-5  
Sequence 5, Application US/07903029  
Patent No. 5969097  
GENERAL INFORMATION:  
APPLICANT: Wiegand, Roger C.  
APPLICANT: Currie, Mark C.  
APPLICANT: Fok, Kam F.  
TITLE OF INVENTION: Human Guanylin  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Dennis P. Bennett, Monsanto Co., A35G  
STREET: 800 N. Lindbergh Blvd.  
CITY: St. Louis  
STATE: Missouri  
COUNTRY: USA  
ZIP: 63167  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/903,029  
FILING DATE: 19920623  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: Bennett, Dennis A.  
REGISTRATION NUMBER: 34,547  
REFERENCE/DOCKET NUMBER: 07-21(872)A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (314)694-5402  
TELEFAX: (314)694-9209  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 115 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-07-903-029-5

Query Match 87.8%; Score 36; DB 2; Length 115;  
Best Local Similarity 87.5%; Pred. No. 6.5;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGVLYVDG 8  
||| |||  
DB 20 GGVTVVDG 27

RESULT 3  
US-08-859-931A-4  
Sequence 4, Application US/08859931A  
Patent No. 5945510  
GENERAL INFORMATION:  
APPLICANT: FASANO, Alessio  
TITLE OF INVENTION: SUBSTANTIALLY PURE ZONULIN, A  
TITLE OF INVENTION: PHYSIOLOGICAL MODULATOR OF  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SUGHRUE, MIOM, ZINN, MACPEAK & SEAS  
STREET: 2100 Pennsylvania Avenue, N.W., Suite 800  
CITY: Washington, D.C.  
STATE: D.C.  
COUNTRY: U.S.A.  
ZIP: 20037  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/859,931A  
FILING DATE: 21 MAY 1997  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: KIT, Gordon  
REGISTRATION NUMBER: 30,764  
REFERENCE/DOCKET NUMBER: A-6901  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 293-7860  
TELEFAX: (202) 293-7860  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 20 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-859-931A-4

Query Match 82.9%; Score 34; DB 2; Length 20;  
Best Local Similarity 87.5%; Pred. No. 2.4;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGVLYVDG 8  
||| |||  
DB 8 GGVLYVDG 15

RESULT 4  
US-08-093-372-4  
Sequence 4, Application US/08093372  
Patent No. 5530187  
GENERAL INFORMATION:  
APPLICANT: Lamb, Christopher J.  
APPLICANT: Zhu, Qun  
APPLICANT: Maher, Eileen A.  
APPLICANT: Dixon, Richard A.  
TITLE OF INVENTION: TRANSGENIC PLANTS CONTAINING MULTIPLE  
TITLE OF INVENTION: DISEASE RESISTANCE GENES  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pretty, Schroeder, Brueggemann & Clark  
STREET: 444 South Flower Street, Suite 2000  
CITY: Los Angeles  
STATE: CA  
COUNTRY: U.S.A.  
ZIP: 90071-2921  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/093,372  
FILING DATE: 16-JUL-1993  
CLASSIFICATION: 800  
ATTORNEY/AGENT INFORMATION:  
NAME: Reiter, Stephen E.  
REGISTRATION NUMBER: 31,192  
REFERENCE/DOCKET NUMBER: P41 9391  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 619-546-4737  
TELEFAX: 619-546-9392  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 368 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-093-372-4

Query Match 80.5%; Score 33; DB 1; Length 368;  
Best Local Similarity 85.7%; Pred. No. 78;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 GGVVODG 8  
11:11111  
DB 234 GGVVODG 240

## RESULT 5

US-08-741-3327E-11  
; Sequence 11, Application US/08741327E  
; Patent No. 5925354  
; GENERAL INFORMATION:  
; APPLICANT: Fuller, et al.  
; TITLE OF INVENTION: Riboflavin mutants as vaccines  
; TITLE OF INVENTION: against Actinobacillus pleuropneumoniae  
; NUMBER OF SEQUENCES: 17  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: G. Kenneth Smith  
; STREET: 300 S. Wacker Drive  
; CITY: Chicago  
; STATE: IL  
; COUNTRY: USA  
; ZIP: 60606  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: ASCII  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/741.327E  
; FILING DATE: October 28, 1996  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: G. Kenneth Smith  
; REGISTRATION NUMBER: 43,135  
; REFERENCE/DOCKET NUMBER: 97704-E  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312-913-0001  
; TELEFAX: 312-913-9808  
; INFORMATION FOR SEQ ID NO: 11:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 401 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; US-08-741-3327E-11

Query Match 78.0%; Score 32; DB 2; Length 401;  
Best Local Similarity 75.0%; Pred. No. 1.3e+02;  
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGVVODG 8  
11111:1  
DB 137 GGVVVRNG 144

## RESULT 6

US-08-545-809A-125  
; Sequence 125, Application US/08545809A  
; Patent No. 6096878  
; GENERAL INFORMATION:  
; APPLICANT: Honjo, Tasuku  
; APPLICANT: Matsuda, Fumihiko  
; TITLE OF INVENTION: HUMAN IMMUNOGLOBULIN VH GENE  
; TITLE OF INVENTION: SEGMENTS AND DNA FRAGMENTS CONTAINING THE SAME  
; NUMBER OF SEQUENCES: 145  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: Fish & Richardson, P.C.  
; STREET: 225 Franklin Street

CITY: Boston  
STATE: MA  
COUNTRY: US  
ZIP: 02110-2804  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM compatible  
OPERATING SYSTEM: Windows95  
SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/545,809A  
FILING DATE: 27-MAR-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/JP93/00603  
FILING DATE: 10-MAY-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Freeman, John W.  
REGISTRATION NUMBER: 29,066  
REFERENCE/DOCKET NUMBER: 06501/004001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-542-5070  
TELEFAX: 617-542-8906  
TELEX: 200154

INFORMATION FOR SEQ ID NO: 125:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 118 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-545-809A-125

Query Match 75.6%; Score 31; DB 3; Length 118;  
Best Local Similarity 75.0%; Pred. No. 54;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGVVODG 8  
111:111  
DB 27 GGVVODG 34

## RESULT 7

US-08-716-301-6  
; Sequence 6, Application US/08716301  
; Patent No. 5821090  
; GENERAL INFORMATION:  
; APPLICANT: Doval, Jose Luis Revuelta  
; APPLICANT: Serna, Maria Jose Butiago  
; TITLE OF INVENTION: Riboflavin biosynthesis in fungi  
; NUMBER OF SEQUENCES: 12  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: Kell & Weinkauff  
; STREET: 1101 Connecticut Avenue  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20036  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb storage  
; COMPUTER: IBM compatible  
; OPERATING SYSTEM: MS-DOS version 7.0  
; SOFTWARE: WordPerfect version 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/716,301  
; FILING DATE: 24-SEP-1996  
; CLASSIFICATION: 435  
; CLASSIFICATION: C 12 N 15/53  
; CLASSIFICATION: C 12 N 15/54  
; CLASSIFICATION: C 12 N 15/55  
; CLASSIFICATION: C 12 N 15/81  
; CLASSIFICATION: C 12 N 1/19  
; CLASSIFICATION: C 12 P 25/00

CLASSIFICATION: C 12 R 1:865  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/EP95/00958  
FILING DATE: 15-MAR-1995  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 212 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-716-301-6

Query Match 75.6%; Score 31; DB 2; Length 212;  
Best Local Similarity 75.0%; Pred. No. 1e+02;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGVLVODG 8  
||| :||  
DB 142 GGVLEKDG 149

RESULT 8  
US-08-774-169-3  
Sequence 3, Application US/08774169  
Patent No. 5756332  
GENERAL INFORMATION:  
APPLICANT: Hillman, Jennifer L.  
TITLE OF INVENTION: A NOVEL GUANOSINE MONOPHOSPHATE REDUCTASE  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/774,169  
FILING DATE: Herewith  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0182 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166  
TELEX:  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 346 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: GenBank  
CLONE: 473772  
US-08-774-169-3

Query Match 75.6%; Score 31; DB 1; Length 346;  
Best Local Similarity 67.5%; Pred. No. 1.7e+02;  
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGVLVODG 8  
||| :||  
DB 213 GGMIVSDG 220

RESULT 9  
US-07-903-029-6  
Sequence 6, Application US/07903029  
Patent No. 5969097  
GENERAL INFORMATION:  
APPLICANT: Wiegand, Roger C.  
APPLICANT: Currie, Mark C.  
TITLE OF INVENTION: Human Guanylin  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Dennis A. Bennett, Monsanto Co., 135G  
STREET: 800 N. Lindbergh Blvd.  
CITY: St. Louis  
STATE: Missouri  
COUNTRY: USA  
ZIP: 63167  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/903,029  
FILING DATE: 19920623  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: Bennett, Dennis A.  
REGISTRATION NUMBER: 34,547  
REFERENCE/DOCKET NUMBER: 07-21(872)A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (314)694-5402  
TELEFAX: (314)694-9009  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 115 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-07-903-029-6

Query Match 73.2%; Score 30; DB 2; Length 115;  
Best Local Similarity 85.7%; Pred. No. 80;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GVLVODG 8  
||| :||  
DB 21 GVTVODG 27

RESULT 10  
US-08-545-809A-108  
Sequence 108, Application US/08545809A  
Patent No. 6096878  
GENERAL INFORMATION:  
APPLICANT: Honjo, Tasuku  
APPLICANT: Matsuda, Fumihiko  
TITLE OF INVENTION: HUMAN IMMUNOGLOBULIN VH GENE  
SEGMENTS AND DNA FRAGMENTS CONTAINING THE SAME  
NUMBER OF SEQUENCES: 145  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson, P.C.  
STREET: 225 Franklin Street  
CITY: Boston  
STATE: MA  
COUNTRY: US  
ZIP: 02110-2804

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: Windows95  
SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/545,809A  
FILING DATE: 27-MAR-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/JP93/00603  
FILING DATE: 10-MAY-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Freeman, John W.  
REGISTRATION NUMBER: 29,066  
REFERENCE/DOCKET NUMBER: 06501/004001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-542-5070  
TELEFAX: 617-542-8906  
TELEX: 200154  
INFORMATION FOR SEQ ID NO: 108:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 118 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-545-809A-108

Query Match 73.2%; Score 30; DB 3; Length 118;  
Best Local Similarity 75.0%; Pred. No. 83;  
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GGVLVODG 8  
11111111  
DB 27 GGALVDPG 34

RESULT 11  
US-08-475-000-16  
Sequence 16, Application US/084755000  
Patent No. 5811267  
GENERAL INFORMATION:  
APPLICANT: RING, DAVID B.  
TITLE OF INVENTION: ANTIGEN-BINDING SITES OF ANTIBODY  
TITLE OF INVENTION: MOLECULES SPECIFIC FOR CANCER ANTIGENS  
NUMBER OF SEQUENCES: 18  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CHIRON CORPORATION  
STREET: INTELLECTUAL PROPERTY R440 - P.O. BOX 8097  
CITY: Emeryville  
STATE: CA  
COUNTRY: USA  
ZIP: 94662-8097  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/475,000  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: SAVERIDE, PAUL  
REGISTRATION NUMBER: 36,914  
REFERENCE/DOCKET NUMBER: 0850,007  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (510) 601-2585  
TELEFAX: (510) 655-3542  
INFORMATION FOR SEQ ID NO: 16:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 119 amino acids  
TYPE: amino acid

TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-475-000-16

Query Match 73.2%; Score 30; DB 2; Length 119;  
Best Local Similarity 75.0%; Pred. No. 83;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGVLVODG 8  
11111111  
DB 8 GGVLVRPG 15

RESULT 12  
US-08-483-199-16  
Sequence 16, Application US/08483199  
Patent No. 5849877  
GENERAL INFORMATION:  
APPLICANT: RING, DAVID B.  
TITLE OF INVENTION: ANTIGEN-BINDING SITES OF ANTIBODY  
TITLE OF INVENTION: MOLECULES SPECIFIC FOR CANCER ANTIGENS  
NUMBER OF SEQUENCES: 18  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CHIRON CORPORATION  
STREET: INTELLECTUAL PROPERTY - R440, P.O. BOX 8097  
CITY: Emeryville  
STATE: CA  
COUNTRY: USA  
ZIP: 94662-8097  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/483,199  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: SAVERIDE, PAUL  
REGISTRATION NUMBER: 36,914  
REFERENCE/DOCKET NUMBER: 0850,009  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (510) 601-2585  
TELEFAX: (510) 655-3542  
INFORMATION FOR SEQ ID NO: 16:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 119 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-483-199-16

Query Match 73.2%; Score 30; DB 2; Length 119;  
Best Local Similarity 75.0%; Pred. No. 83;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGVLVODG 8  
11111111  
DB 8 GGVLVRPG 15

RESULT 13  
US-08-484-508-16  
Sequence 16, Application US/08484508  
Patent No. 5948647  
GENERAL INFORMATION:  
APPLICANT: RING, DAVID B.  
TITLE OF INVENTION: ANTIGEN-BINDING SITES OF ANTIBODY  
TITLE OF INVENTION: MOLECULES SPECIFIC FOR CANCER ANTIGENS  
NUMBER OF SEQUENCES: 18

CORRESPONDENCE ADDRESS:  
ADDRESSEE: CHIRON CORPORATION  
STREET: INTELLECTUAL PROPERTY - R440, P.O. BOX 8097  
CITY: Emeryville  
STATE: CA  
COUNTRY: USA  
ZIP: 94662-8097  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/484,508  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:  
NAME: SAVERIDE, PAUL  
REGISTRATION NUMBER: 36,914  
REFERENCE/DOCKET NUMBER: 0850,008  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (510) 601-585  
TELEFAX: (510) 655-3542  
INFORMATION FOR SEQ ID NO: 16:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 119 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-484-508-16

Query Match 73.28; Score 30; DB 2; Length 119;  
Best Local Similarity 73.08; Pred. No. 83;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGVLYVDG 8  
| | | | |  
DB 8 GGVLYVRG 15

RESULT 14  
US-08-428-197-38  
Sequence 38, Application US/08428197  
Patent No. 5891438  
GENERAL INFORMATION:  
APPLICANT: SILVERMAN, GREGG J.  
TITLE OF INVENTION: METHOD FOR STIMULATING PRODUCTION OF  
TITLE OF INVENTION: VARIABLE REGION GENE FAMILY RESTRICTED ANTIBODIES THROUGH  
TITLE OF INVENTION: VACCINATION WITH A B-CELL SUPERANTIGEN AND CONJUGATES  
NUMBER OF SEQUENCES: 51  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Spensley Horn Jubas & Lubitz  
STREET: 1880 Century Park East - Suite 500  
CITY: Los Angeles  
STATE: California  
COUNTRY: USA  
ZIP: 90067  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/428,197  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA: PCT/US93/10555  
APPLICATION NUMBER: 29-OCT-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Howells, Stacy L.

REGISTRATION NUMBER: 34,842  
REFERENCE/DOCKET NUMBER: FD-2630  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 455-5100  
TELEFAX: (619) 455-5110  
INFORMATION FOR SEQ ID NO: 38:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 123 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-428-197-38

Query Match 73.28; Score 30; DB 2; Length 123;  
Best Local Similarity 75.08; Pred. No. 86;  
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GGVLYVDG 8  
| | | | |  
DB 8 GGVLYVRG 15

RESULT 15  
PCT-US93-10555-38  
Sequence 38, Application PC/TUS9310555  
GENERAL INFORMATION:  
APPLICANT: SILVERMAN, GREGG J.  
TITLE OF INVENTION: METHOD FOR STIMULATING PRODUCTION OF  
TITLE OF INVENTION: VARIABLE REGION GENE FAMILY RESTRICTED ANTIBODIES THROUGH  
TITLE OF INVENTION: VACCINATION WITH A B-CELL SUPERANTIGEN AND CONJUGATES  
NUMBER OF SEQUENCES: 51  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Spensley Horn Jubas & Lubitz  
STREET: 1880 Century Park East - Suite 500  
CITY: Los Angeles  
STATE: California  
COUNTRY: USA  
ZIP: 90067  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US93/10555  
FILING DATE: 29-OCT-1993  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Howells, Stacy L.  
REGISTRATION NUMBER: 34,842  
REFERENCE/DOCKET NUMBER: FD-2630  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 455-5100  
TELEFAX: (619) 455-5110  
INFORMATION FOR SEQ ID NO: 38:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 123 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
PCT-US93-10555-38

Query Match 73.28; Score 30; DB 5; Length 123;  
Best Local Similarity 75.08; Pred. No. 86;  
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GGVLYVDG 8  
| | | | |  
DB 8 GGVLYVRG 15



Search completed: June 13, 2001, 14:16:36  
Job time: 497 sec

---

*This Page Blank (uspto)*

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 13, 2001, 14:14:39 ; Search time 229.28 Seconds  
(without alignments)  
1.995 Million cell updates/sec

Title: PCT-US01-05825a-17

Perfect score: 47

Sequence: 1 GGLCVQPG 8

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 390729 seqs, 57163235 residues

Total number of hits satisfying chosen parameters: 390729

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A\_Geneseq\_0401.\*  
1: /SID56/gcgdata/geneseq/geneseq/AA1980.DAT.\*  
2: /SID56/gcgdata/geneseq/geneseq/AA1981.DAT.\*  
3: /SID56/gcgdata/geneseq/geneseq/AA1982.DAT.\*  
4: /SID56/gcgdata/geneseq/geneseq/AA1983.DAT.\*  
5: /SID56/gcgdata/geneseq/geneseq/AA1984.DAT.\*  
6: /SID56/gcgdata/geneseq/geneseq/AA1985.DAT.\*  
7: /SID56/gcgdata/geneseq/geneseq/AA1986.DAT.\*  
8: /SID56/gcgdata/geneseq/geneseq/AA1987.DAT.\*  
9: /SID56/gcgdata/geneseq/geneseq/AA1988.DAT.\*  
10: /SID56/gcgdata/geneseq/geneseq/AA1989.DAT.\*  
11: /SID56/gcgdata/geneseq/geneseq/AA1990.DAT.\*  
12: /SID56/gcgdata/geneseq/geneseq/AA1991.DAT.\*  
13: /SID56/gcgdata/geneseq/geneseq/AA1992.DAT.\*  
14: /SID56/gcgdata/geneseq/geneseq/AA1993.DAT.\*  
15: /SID56/gcgdata/geneseq/geneseq/AA1994.DAT.\*  
16: /SID56/gcgdata/geneseq/geneseq/AA1995.DAT.\*  
17: /SID56/gcgdata/geneseq/geneseq/AA1996.DAT.\*  
18: /SID56/gcgdata/geneseq/geneseq/AA1997.DAT.\*  
19: /SID56/gcgdata/geneseq/geneseq/AA1998.DAT.\*  
20: /SID56/gcgdata/geneseq/geneseq/AA1999.DAT.\*  
21: /SID56/gcgdata/geneseq/geneseq/AA2000.DAT.\*  
22: /SID56/gcgdata/geneseq/geneseq/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	47	100.0	8	21	Y79121
2	44	93.6	8	21	Y79117
3	41	87.2	31	20	W88384
4	41	87.2	73	20	W88389
5	41	87.2	115	21	B41718
6	41	87.2	153	21	B41638
7	41	87.2	158	20	W88388
8	41	87.2	169	20	W88390
9	41	87.2	224	20	Y59870
10	41	87.2	247	21	Y52139
11	41	87.2	254	20	W88382

12	41	87.2	265	21	B42204	Human OREF ORF1968
13	41	87.2	273	20	Y41769	Human PRO213-1 pro
14	41	87.2	273	20	Y41770	Human PRO1330 prot
15	41	87.2	273	20	Y41771	Human PRO1449 prot
16	41	87.2	273	20	W88381	Human neuro-growth
17	41	87.2	273	21	B44325	Human PRO213-1 pro
18	41	87.2	273	21	B44326	Human PRO1330 prot
19	41	87.2	273	21	B44327	Human PRO1449 prot
20	41	87.2	273	21	B18673	Amino acid sequenc
21	41	87.2	273	21	B18674	Amino acid sequenc
22	41	87.2	273	21	B18675	Amino acid sequenc
23	41	87.2	273	21	B24042	Human PRO213 prote
24	41	87.2	273	21	B24043	Human PRO1330 prot
25	41	87.2	273	21	B24044	Human PRO1449 prot
26	41	87.2	273	21	B01376	Neuron-associated
27	41	87.2	273	21	Y52137	Human TANGO 125 (T
28	41	87.2	295	20	Y41685	Human PRO213 prote
29	41	87.2	295	21	B44241	Human PRO213 (UNQ1
30	41	87.2	307	21	B41644	Human OREF ORF1408
31	41	87.2	379	20	Y08065	Human EGF-like hom
32	41	87.2	379	20	Y13345	Amino acid sequenc
33	41	87.2	379	21	B24397	Human PRO217 prote
34	41	87.2	379	21	Y70659	Human PRO217 prote
35	41	87.2	379	21	Y44822	Human molecule ass
36	41	87.2	379	22	B53076	Human angiogenesis
37	41	87.2	380	20	W88726	Secreted protein e
38	39	83.0	8	21	Y79109	Peptide antagonist
39	39	83.0	8	21	Y79122	Peptide antagonist
40	39	83.0	8	21	Y79125	Peptide antagonist
41	37	78.7	8	21	Y79123	Peptide antagonist
42	36	76.6	8	21	Y79105	Peptide antagonist
43	36	76.6	8	21	Y79118	Peptide antagonist
44	36	76.6	255	20	W88397	Mouse neuro-growth
45	36	76.6	275	21	Y52141	Mouse TANGO 125 (T

#### ALIGNMENTS

RESULT	1
ID	Y79121 standard; Peptide: 8 AA.
XX	Y79121;
AC	Y79121;
XX	
DT	05-JUN-2000 (first entry)
XX	
DE	Peptide antagonist of zonulin.
XX	
KW	zonulin; antagonist; zonula occludens toxin receptor;
KW	blood-brain barrier; antiinflammatory; cerebroprotective;
KW	neuroprotective; dermatological; antitumor; antiviral;
KW	antibacterial; cytostatic; anti-HIV; vulnerary; antiallergic;
KW	hypotensive; immunosuppressive; antiparasitic; vasotropic;
KW	gastrointestinal inflammation; therapy.
OS	Synthetic.
XX	
PN	WO200007609-A1.
XX	
PD	17-FEB-2000.
XX	
PF	28-JUL-1999; 99MO-US16683.
XX	
PR	03-AUG-1998; 98US-0127815.
XX	
PA	(UYWA-) UNIV MARYLAND BALTIMORE.
XX	
PI	Fasano A;
XX	
DR	WPI; 2000-205565/18.
XX	
PT	New peptide antagonist of zonulin useful as antiinflammatory agent for

PT treating cerebral ischemia, stroke, cerebral edema, gastritis,  
XX shigellosis, viral gastroenteritis, meningitis, encephalomyelitis  
XX  
PS Claim 1; Page 46; 69pp; English.

CC This present sequence is that of a peptide antagonist of zonulin  
CC (2), one of 25 such peptides (see 79105-29) of the invention,  
CC which bind to a zonula occludens toxin (ZOT) receptor, yet do not  
CC physiologically modulate the opening of mammalian tight junctions  
CC (TJ). The peptide antagonists are based on a common motif of ZOT  
CC and human zonulins, which is believed to be critical for receptor  
CC binding. They can be prepared by chemical synthesis or by use of  
CC recombinant DNA techniques. The peptide antagonists are used as an  
CC antiinflammatory agents in the treatment of gastrointestinal  
CC inflammation, where they bind to the ZOT receptor in the intestine  
CC and yet does not physiologically modulate the opening of TJ in the  
CC intestine. Gastrointestinal inflammation conditions give rise to  
CC increased intestinal permeability and the peptide is useful for  
CC treating intestinal conditions that cause protein losing enteropathy  
CC caused by infection, e.g. Clostridium difficile infection,  
CC enterocolitis, shigellosis, viral gastroenteritis, parasite  
CC mucosal erosion or ulcerations, e.g. gastritis, gastric cancer,  
CC collagenous colitis, inflammatory bowel disease, diseases with  
CC lymphatic obstruction, e.g. congenital intestinal lymphangiectasia,  
CC sarcoïdosis lymphoma, mesenteric tuberculosis, and after surgical  
CC correction of congenital heart disease with Fontan's operation,  
CC mucosal diseases without ulceration, e.g. Menetrier's disease,  
CC coeliac disease, eosinophilic gastroenteritis, and immune diseases,  
CC e.g. systemic lupus erythematosus or food allergies, primarily to  
CC milk.  
XX  
SQ Sequence 8 AA:

Query Match 100.0%; Score 47; DB 21; Length 8;  
Best Local Similarity 100.0%; Pred. No. 3.2e+05;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GGLCVQPG 8  
|||  
1 gglcvqpg 8

RESULT 2  
79117  
ID 79117 standard; Peptide: 8 AA.

AC 79117;

DT 05-JUN-2000 (first entry)

XX Peptide antagonist of zonulin.

XX Zonulin; antagonist; zonula occludens toxin receptor;  
XX blood-brain barrier; antiinflammatory; cerebroprotective;  
XX neuroprotective; dermatological; antitumor; antiviral;  
XX antibacterial; cytosstatic; anti-HIV; vulnerary; antiallergic;  
XX hypotensive; immunosuppressive; antiparasitic; vasotropic;  
XX gastrointestinal inflammation; therapy.

OS Synthetic.

PN WO200007609-A1.

XX 17-FEB-2000.

XX 28-JUL-1999; 99WO-US16683.

XX 03-AUG-1998; 98US-0127815.

XX (UYMA-) UNIV MARYLAND BALTIMORE.

PI Fasano A;  
XX  
DR WPI: 2000-205565/18.  
XX

PT New peptide antagonist of zonulin useful as antiinflammatory agent for  
PT treating cerebral ischemia, stroke, cerebral edema, gastritis,  
PT shigellosis, viral gastroenteritis, meningitis, encephalomyelitis  
XX  
PS Claim 1; Page 45; 69pp; English.

CC This present sequence is that of a peptide antagonist of zonulin  
CC (2), one of 25 such peptides (see 79105-29) of the invention,  
CC which bind to a zonula occludens toxin (ZOT) receptor, yet do not  
CC physiologically modulate the opening of mammalian tight junctions  
CC (TJ). The peptide antagonists are based on a common motif of ZOT  
CC and human zonulins, which is believed to be critical for receptor  
CC binding. They can be prepared by chemical synthesis or by use of  
CC recombinant DNA techniques. The peptide antagonists are used as an  
CC antiinflammatory agents in the treatment of gastrointestinal  
CC inflammation, where they bind to the ZOT receptor in the intestine  
CC and yet does not physiologically modulate the opening of TJ in the  
CC intestine. Gastrointestinal inflammation conditions give rise to  
CC increased intestinal permeability and the peptide is useful for  
CC treating intestinal conditions that cause protein losing enteropathy  
CC caused by infection, e.g. Clostridium difficile infection,  
CC enterocolitis, shigellosis, viral gastroenteritis, parasite  
CC mucosal erosion or ulcerations, e.g. gastritis, gastric cancer,  
CC collagenous colitis, inflammatory bowel disease, diseases marked by  
CC lymphatic obstruction, e.g. congenital intestinal lymphangiectasia,  
CC sarcoïdosis lymphoma, mesenteric tuberculosis, and after surgical  
CC correction of congenital heart disease with Fontan's operation,  
CC mucosal diseases without ulceration, e.g. Menetrier's disease,  
CC coeliac disease, eosinophilic gastroenteritis, and immune diseases,  
CC e.g. systemic lupus erythematosus or food allergies, primarily to  
CC milk.  
XX  
SQ Sequence 8 AA:

Query Match 93.6%; Score 44; DB 21; Length 8;  
Best Local Similarity 87.5%; Pred. No. 3.2e+05;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GGLCVQPG 8  
|||  
1 gglcvqpg 8

RESULT 3  
W88384  
ID W88384 standard; Peptide: 31 AA.

AC W88384;

DT 26-APR-1999 (first entry)

XX Human Zneul EGF-like domain 1.

XX Zneul-1; neuro-growth factor-like protein; human; breast cancer;  
XX glioblastoma; pituitary adenoma; Alzheimer's disease; therapy;  
XX nerve regeneration; haematopoiesis; fertility; contraception;  
XX antibody; epidermal growth factor; EGF.

XX Homo sapiens.

PN WO9857983-A2.

XX 23-DEC-1998.

XX 18-JUN-1998; 98WO-US12763.

XX 18-JUN-1997; 97US-0878322.

PR	18-JUN-1997;	97US-0050143.
XX		
PA	( ZYMO ) ZYMOGENETICS INC.	
XX		
P1	Blumberg H, Jelinek LJ, Lehner JM, Sheppard PO;	
P1	Whitmore TE;	
XX		
DR	WPI: 1999-095324/08.	
XX		
PT	New mammalian Zneul polypeptides - used to, e.g. treat Alzheimer's	
PT	disease, cancer and to repopulate blood cells	
XX		
PS	Claim 6; Page 53; 70pp; English.	
XX		
CC	This peptide comprises the first epidermal growth factor-like	
CC	domain (EGF-1), i.e. amino acid residues 105-135, of human Zneul	
CC	(see W88381), a new neuro-growth factor-like protein. Zneul can be	
CC	used as a growth, maintenance, or differentiation factor in the	
CC	spinal cord, heart, spleen, testis, thymoid and lymph nodes. It	
CC	may also play a role in breast cancer, glioblastomas, and pituitary	
CC	adenomas. Zneul may also be used to treat Alzheimer's disease,	
CC	cancer, to repopulate blood cells after chemotherapy, to stimulate	
CC	myofibroblast proliferation, stimulate or inhibit growth factors	
CC	made in the placenta, in fertility and contraception, or to	
CC	regenerate nerves. Claimed Zneul polypeptides (see also W88382-97),	
CC	including specific domains of Zneul and epitope-bearing portions of	
CC	Zneul, can be used to raise specific antibodies for use e.g. in	
CC	diagnostic assays.	
XX		
50	Sequence 31 AA:	

```

Query Match      87.2%; Score 41; DB 20; Length 31;
Best Local Similarity 87.5%; Pred No. 2.7;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY      1 GGLCVQPG 8
        |||
        10 ggsccvqp 17

DB

RESULT      4
W88389
ID      W88389 standard; Protein; 73 AA.
XX
XX      W88389;
AC
DT      26-APR-1999 (first entry)
XX
XX      Human zneul partial polypeptide.
DE
XX
KW      zneu-1; neuro-growth factor-like protein; human; breast cancer;
KW      glioblastoma; pituitary adenoma; Alzheimer's disease; therapy;
KW      nerve regeneration; haematopoiesis; fertility; contraception;
KW      antibody.
XX
OS      Homo sapiens.
XX
PN      W09857983-A2.
XX
PD      23-DEC-1998.
XX
XX      18-JUN-1998; 98WO-US12763.
XX
XX      18-JUN-1997; 97US-0878322.
XX      18-JUN-1997; 97US-0050143.
XX
XX      (ZYMO ) ZYMOGENETICS INC.
XX
XX      Blumberg H, Jelinek LJ, Lehner JM, Sheppard PO;
XX      Whitmore TE;
XX      WPI; 1999-095324/08.
DR

```

Query Match	Best Local Similarity	Score 41;	DB 20;	Length 73;
Matches 7;	Conservative 0;	Mismatches 1;	Indels 0;	Gaps 0;
QY	1 GGLCVQPG 8			
DB	10 ggcwvpg 17			
RESULT 5				
ID	B41718 standard; Protein; 115 AA.			
XX	B41718			
AC	B41718;			
XX	08-FEB-2001 (first entry)			
XX				
XX	Human ORFX ORF1482 polypeptide sequence SEQ ID NO:2964.			
XX				
XX	Human: open reading frame; ORFX; detection: cytosolic; hepatotropic;			
KM	vulnerary; antiporiatic; antiparkinsonian; nootropic; neuroprotective;			
KM	anticonvulsant; osteopathic; antiarthritic; immunosuppressant; ceridant;			
KM	immunostimulant; thrombolytic; coagulant; vasotropic; antididiabetic;			
KM	hypotensive; dermatological; immunosuppressive; antiinflammatory;			
KM	antiviral; antibacterial; antifungal; antirheumatic; antithyroid;			
KM	antianemic; gene therapy; cancer; proliferative disorder; hypertension;			
KM	neurodegenerative disorder; osteoarthritis; graft vs host disease;			
KM	cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;			
KM	cholesterol ester storage; systemic lupus erythematosus; infection;			
KM	severe combined immunodeficiency; malaria; autoimmune disorder; asthma;			
KM	allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;			
KM	bone damage; cartilage damage; antiinflammatory disease; coagulation;			
KM	thrombosis; contraceptive.			
XX				
OS	Homo sapiens.			
XX				
PN	WO200058473-A2.			
XX				
PD	05-OCT-2000.			
XX				
PF	31-MAR-2000; 2000WO-US08621.			
XX				
PR	31-MAR-1999; 99US-0127607.			
PR	02-APR-1999; 99US-0127636.			
PR	05-APR-1999; 99US-0127728.			
PR	30-MAR-2000; 2000US-0540763.			
XX				
PA	(CURA-) CURAGEN CORP.			

XX Shimkets RA, Leach M;  
 XX MPI: 2000-602362/57.  
 DR N-PSDB: C75927.  
 XX  
 PT Novel nucleic acids and peptides derived from open reading frame X,  
 PT useful for treating e.g. cancers, proliferative disorders,  
 PT neurodegenerative disorders and cardiovascular disease -  
 XX  
 PS Claim 11; Page 2200; 5507pp; English.  
 CC C74446 to C77606 encode the proteins given in B40237 to B43397, which  
 CC represent the human ORFX open reading frames 1 to 3161. The ORFX  
 CC sequences have activities such as: cytostatic; hepatotropic; vulnery;  
 CC antiproliferative; antiparkinsonian; neurotrophic; neuroprotective; osteopathic;  
 CC anticonvulsant; antiarthritic; immunosuppressant; immunostimulant;  
 CC cardiatic; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive;  
 CC dermatological; immunosuppressive; antineoplastic; antibacterial;  
 CC antiviral; antifungal; antineumatic; antihypertensive; antianemic.  
 CC sequences can be used for determining the presence of or predisposition  
 CC to, or preventing or treating pathological conditions associated with an  
 CC ORFX-associated disorder. The nucleic acids can be used to express ORFX  
 CC proteins in gene therapy vectors. The proteins and nucleic acids may be  
 CC used to treat cancers, proliferative disorders, neurodegenerative  
 CC disorders, osteoarthritis, graft vs host disease, cardiovascular disease,  
 CC diabetes mellitus, hypertension, hypothyroidism, cholesterol ester  
 CC storage, systemic lupus erythematosus, severe combined immunodeficiency  
 CC (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune  
 CC disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and  
 CC cartilage damage, nocturnal haemoglobinuria, antiinflammatory disease; to  
 CC enhance coagulation; to inhibit thrombosis; and as a contraceptive.  
 XX  
 SQ Sequence 115 AA;  
 Query Match 87.2%; Score 41; DB 21; Length 115;  
 Best Local Similarity 87.5%; Pred. No. 8.5;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 OY 1 GGLCVOPG 8  
 XX 111111  
 DB 19 gscvcpq 26  
 RESULT 6  
 ID B41638  
 AC B41638 standard; Protein: 153 AA.  
 XX  
 AC B41638;  
 XX  
 DT 08-FEB-2001 (first entry)  
 XX  
 DE Human ORFX ORF1402 polypeptide sequence SEQ ID NO:2804.  
 XX  
 KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;  
 KW vulnery; antiproliferative; antiparkinsonian; neurotrophic; neuroprotective;  
 KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiatic;  
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;  
 KW hypotensive; dermatological; immunosuppressive; antineoplastic;  
 KW antiviral; antibacterial; antineumatic; antihypertensive; antianemic;  
 KW antianemic; gene therapy; cancer; proliferative disorder; hypertension;  
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;  
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;  
 KW cholesterol ester storage; systemic lupus erythematosus; infection;  
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;  
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;  
 KW bone damage; cartilage damage; antiinflammatory disease; coagulation;  
 KW thrombosis; contraceptive.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200058473-A2.

XX  
 PD 05-OCT-2000.  
 XX  
 PF 31-MAR-2000; 2000WO-US08621.  
 XX  
 PR 31-MAR-1999; 99US-0127607.  
 PR 02-APR-1999; 99US-0127636.  
 PR 05-APR-1999; 99US-0127728.  
 PR 30-MAR-2000; 2000US-0540763.  
 XX  
 PA (CURA-) CURAGEN CORP.  
 XX  
 PI Shimkets RA, Leach M;  
 XX MPI: 2000-602362/57.  
 DR N-PSDB: C75847.  
 XX  
 PT Novel nucleic acids and peptides derived from open reading frame X,  
 PT useful for treating e.g. cancers, proliferative disorders,  
 PT neurodegenerative disorders and cardiovascular disease -  
 XX  
 PS Claim 11; Page 2044; 5507pp; English.  
 CC C74446 to C77606 encode the proteins given in B40237 to B43397, which  
 CC represent the human ORFX open reading frames 1 to 3161. The ORFX  
 CC sequences have activities such as: cytostatic; hepatotropic; vulnery;  
 CC antiproliferative; antiparkinsonian; neurotrophic; neuroprotective; osteopathic;  
 CC anticonvulsant; antiarthritic; immunosuppressant; immunostimulant;  
 CC cardiatic; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive;  
 CC dermatological; immunosuppressive; antineoplastic; antibacterial;  
 CC antiviral; antifungal; antineumatic; antihypertensive; antianemic.  
 CC sequences can be used for determining the presence of or predisposition  
 CC to, or preventing or treating pathological conditions associated with an  
 CC ORFX-associated disorder. The nucleic acids can be used to express ORFX  
 CC proteins in gene therapy vectors. The proteins and nucleic acids may be  
 CC used to treat cancers, proliferative disorders, neurodegenerative  
 CC disorders, osteoarthritis, graft vs host disease, cardiovascular disease,  
 CC diabetes mellitus, hypertension, hypothyroidism, cholesterol ester  
 CC storage, systemic lupus erythematosus, severe combined immunodeficiency  
 CC (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune  
 CC disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and  
 CC cartilage damage, nocturnal haemoglobinuria, antiinflammatory disease; to  
 CC enhance coagulation; to inhibit thrombosis; and as a contraceptive.  
 XX  
 SQ Sequence 153 AA;  
 Query Match 87.2%; Score 41; DB 21; Length 153;  
 Best Local Similarity 87.5%; Pred. No. 11;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 OY 1 GGLCVOPG 8  
 XX 111111  
 DB 91 gscvcpq 98  
 RESULT 7  
 ID W88388  
 AC W88388 standard; Protein: 158 AA.  
 XX  
 AC W88388;  
 XX  
 DT 26-APR-1999 (first entry)  
 XX  
 DE Human Zneul partial polypeptide.  
 XX  
 KW Zneul-1; neuro-growth factor-like protein; human; breast cancer;  
 KW glioblastoma; pituitary adenoma; Alzheimer's disease; therapy;  
 KW nerve regeneration; haematopoiesis; fertility; contraception;  
 KW antibody.  
 XX  
 OS Homo sapiens.  
 XX  
 PN

PN W09857983-A2.  
XX  
PD 23-DEC-1998.  
XX  
PF 18-JUN-1998; 98WO-US12763.  
XX  
PR 18-JUN-1997; 97US-0878322.  
PR 18-JUN-1997; 97US-0050143.  
XX  
PA (ZYMO ) ZYMOGENETICS INC.  
XX  
PI Blumberg H, Jellinek LJ, Lehner JM, Sheppard PO;  
PI Whitmore TE;  
XX WPI: 1999-095324/08.  
DR  
XX  
PT New mammalian Zneul polypeptides - used to, e.g. treat Alzheimer's  
PT disease, cancer and to repopulate blood cells  
XX  
PS Claim 6; Page 56; 70pp; English.  
XX  
XX This claimed polypeptide comprises amino acid residues 1-158 of the  
CC human Zneul mature protein (see W88382), i.e. the first HSMHC3W5A  
CC homology domain and first EGF-like domain of Zneul. Zneul is a new  
CC neuro-growth factor-like protein that can be used as a growth, heart,  
CC maintenance, or differentiation factor in the spinal cord, heart,  
CC spleen, testis, thyroid and lymph nodes. Zneul may also play a  
CC role in breast cancer, glioblastomas, and pituitary adenomas.  
CC Zneul can be used to treat Alzheimer's disease, cancer, to  
CC repopulate blood cells after chemotherapy, to stimulate  
CC myofibroblast proliferation, stimulate or inhibit growth factors  
CC made in the placenta, in fertility and contraception, or to  
CC regenerate nerves. Claimed Zneul polypeptides (see also W88382-97),  
CC including specific domains of Zneul and epitope-bearing portions of  
CC Zneul, can be used to raise specific antibodies for use e.g. in  
CC diagnostic assays.  
XX  
SQ Sequence 158 AA:

Query Match 87.2%; Score 41; DB 20; Length 158;  
Best Local Similarity 87.5%; Pred. No. 11;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGLCVQPG 8  
DB 95 ggcvcvpg 102  
II IIIII

RESULT 8  
W88390  
ID W88390 standard; Protein: 169 AA.  
XX  
AC W88390;  
XX  
DT 26-APR-1999 (first entry)  
XX  
DE Human Zneul partial polypeptide.  
XX  
XX Zneul-1; neuro-growth factor-like protein; human; breast cancer;  
KW glioblastoma; pituitary adenoma; Alzheimer's disease; therapy;  
KW nerve regeneration; haematopoiesis; fertility; contraception;  
KW antibody.  
XX  
XX Homo sapiens.  
OS  
XX  
PN W09857983-A2.  
XX  
PD 23-DEC-1998.  
XX  
PF 18-JUN-1998; 98WO-US12763.  
PF 18-JUN-1998; 98WO-US12763.  
PR 18-JUN-1997; 97US-0878322.

PR 18-JUN-1997; 97US-0050143.  
XX  
PA (ZYMO ) ZYMOGENETICS INC.  
XX  
PI Blumberg H, Jellinek LJ, Lehner JM, Sheppard PO;  
PI Whitmore TE;  
XX WPI: 1999-095324/08.  
DR  
XX  
PT New mammalian Zneul polypeptides - used to, e.g. treat Alzheimer's  
PT disease, cancer and to repopulate blood cells  
XX  
PS Claim 6; Page 57-58; 70pp; English.  
XX  
XX This claimed polypeptide comprises amino acid residues 105-273 of  
CC the human Zneul mature protein (see W88382), i.e. the 2 epidermal  
CC growth factor-like domains (see also W88384-85) and the second  
CC HSMHC3W5A homology domain of Zneul. Zneul is a new neuro-growth  
CC factor-like protein that can be used as a growth, maintenance, or  
CC differentiation factor in the spinal cord, heart, spleen, testis,  
CC thyroid and lymph nodes. Zneul may also play a role in breast  
CC cancer, glioblastomas, and pituitary adenomas. Zneul can be used  
CC to treat Alzheimer's disease, cancer, to repopulate blood cells  
CC after chemotherapy, to stimulate myofibroblast proliferation,  
CC stimulate or inhibit growth factors made in the placenta, in  
CC fertility and contraception, or to regenerate nerves. Claimed  
CC Zneul polypeptides (see also W88382-97), including specific domains  
CC of Zneul and epitope-bearing portions of Zneul, can be used to  
CC raise specific antibodies for use e.g. in diagnostic assays.  
XX  
SQ Sequence 169 AA:

Query Match 87.2%; Score 41; DB 20; Length 169;  
Best Local Similarity 87.5%; Pred. No. 12;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGLCVQPG 8  
DB 10 ggcvcvpg 17  
II IIIII

RESULT 9  
Y59870  
ID Y59870 standard; Protein: 224 AA.  
XX  
AC Y59870;  
XX  
DT 19-JAN-2000 (first entry)  
XX  
DE Human normal uterus tissue derived protein 33.  
XX  
KW Human; uterus; cancer; treatment; anticancer; cytostatic; gene therapy;  
KW EST; expressed sequence tag..  
XX  
XX Homo sapiens.  
OS  
XX  
PN DE19817946-A1.  
XX  
PD 21-OCT-1999.  
XX  
PF 17-APR-1998; 98DE-1017946.  
PF 17-APR-1998; 98DE-1017946.  
PR 17-APR-1998; 98DE-1017946.  
XX  
XX (METRA-) METAGEN GES GENOMFORSCHUNG MBH.  
XX  
PI Rosenthal A, Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E;  
PI WPI: 1999-591956/51.  
DR N-PSDB: Z41336.  
XX  
PT New nucleic acid sequences expressed in normal uterine tissues, and

PT derived polypeptides, for treatment of uterine cancer and  
 PT identification of therapeutic agents  
 PS Claim 23; Page 134; 154pp; German.  
 XX  
 CC This invention describes novel cDNA sequences (A) highly expressed in  
 CC normal uterine tissue which can have anticancer and cytostatic activity  
 CC and can be used for gene therapy. (A) are used (i) for recombinant  
 CC expression of polypeptides (B) and (ii) to isolate complete genes.  
 CC (B) are used (i) to identify agents suitable for treatment of uterine  
 CC cancer; (ii) directly for treating this form of cancer (including  
 CC expression from gene therapy vectors) and (iii) for generation of  
 CC specific antibodies. (A) are identified by assembling ESTs (expressed  
 CC sequence tags) from a particular tissue type before comparison of  
 CC expression patterns. This allows a significantly longer fragment of the  
 CC gene to be revealed, so should reduce the number of failures associated  
 CC with the fact that ESTs from different libraries may represent  
 CC different parts of the same unknown gene, distorting the estimated  
 CC frequency of occurrence in a particular tissue. Y59838-Y59892 represent  
 CC protein fragments encoded by the human uterine tissue derived cDNA  
 CC fragments represented in Z41325-Z41385.  
 CC  
 SO Sequence 224 AA;  
 Query Match 87.2%; Score 41; DB 20; Length 224;  
 Best Local Similarity 87.5%; Pred. No. 15;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 OY 1 GGLCVOPG 8  
 DB 65 ggcscvqpg 72  
 RESULT 10  
 Y52139  
 ID Y52139 standard; Protein; 247 AA.  
 AC Y52139;  
 XX  
 DT 01-FEB-2000 (first entry)  
 XX  
 DE Human TANGO 125b (T125b) amino acid sequence.  
 XX  
 KW TANGO 125; T125; alternative splice variant; EGF domain; antibody;  
 KW secreted protein; agonist; antagonist; predictive medicine; treatment;  
 KW forensic biology.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO954437-A2.  
 PD  
 XX 28-OCT-1999.  
 PF  
 XX 23-APR-1999; 99WO-US08900.  
 PR  
 XX 23-APR-1998; 98US-0065363.  
 PR 23-APR-1999; 99US-0065363.  
 XX  
 PA (MILL-) MILLENNIUM BIOTHERAPEUTICS INC.  
 XX  
 PI Holtzman DA;  
 DR WPI; 2000-013240/01.  
 DR N-PSDB; Z37133.  
 XX  
 PT Novel polynucleotides and polypeptides used to modulate a variety of  
 PT cellular processes  
 XX  
 PS Example 4; Fig 6; 120pp; English.  
 CC Y52138-Y52140 are amino acid sequences of spliced variants of TANGO 125  
 CC (T125). The T125 protein (Y52137) has two epidermal growth factor

CC (EGF)-like domains at amino acids 107-134 and 141-176 and is predicted to  
 CC have a molecular weight of approximately 30kD. T125 is predicted to have  
 CC no transmembrane domains and appears to be a secreted protein. The  
 CC sequences of all variants of T125 are used in the invention to create  
 CC antibodies which selectively bind to T125. The T125 polypeptide is used  
 CC to modulate a variety of cellular processes. It can be used to produce,  
 CC fusion proteins. The protein may also be used to produce antibodies, and  
 CC to identify T125 antagonists and agonists. The T125 polynucleotides,  
 CC polypeptides, homologues and antibodies can be used in screening assays;  
 CC predictive medicine; and methods of treatment of T125 associated  
 CC disorders. The T125 polynucleotides can be used to express the protein;  
 CC to detect T125 mRNA; to detect genetic alterations in the T125 gene; in  
 CC forensic biology; and as a source of primers and probes.  
 CC  
 SO Sequence 247 AA;  
 Query Match 87.2%; Score 41; DB 21; Length 247;  
 Best Local Similarity 87.5%; Pred. No. 17;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 OY 1 GGLCVOPG 8  
 DB 88 ggcscvqpg 95  
 RESULT 11  
 W88382  
 ID W88382 standard; Protein; 254 AA.  
 AC W88382;  
 XX  
 DT 26-APR-1999 (first entry)  
 XX  
 DE Human neuro-growth factor-like protein Zneu1 mature polypeptide.  
 XX  
 KW Zneu-1; neuro-growth factor-like protein; human; breast cancer;  
 KW glioblastoma; pituitary adenoma; Alzheimer's disease; therapy;  
 KW nerve regeneration; haematopoiesis; fertility; contraception;  
 KW antibody.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key  
 FT Location/Qualifiers  
 FT 1..85  
 FT /note="hydrophilic domain (HSM1), homologous to  
 FT an HSMHC3W5A domain"  
 FT  
 FT Domain  
 FT 86..116  
 FT /note="epidermal growth factor-like domain 1"  
 FT 117..158  
 FT /note="epidermal growth factor-like domain 2"  
 FT 159..254  
 FT /note="domain HSM2 homologous to an HSMHC3W5A  
 FT domain"  
 FT  
 XX  
 PN WO9857983-A2.  
 PD  
 XX 23-DEC-1998.  
 PF  
 XX 18-JUN-1998; 98WO-US12763.  
 PR  
 XX 18-JUN-1997; 97US-0878322.  
 PR 18-JUN-1997; 97US-0050143.  
 XX  
 PA (ZYMO) ZYMOGENETICS INC.  
 XX  
 PI Blumberg H, Jelinek LJ, Lehner JM, Sheppard PO;  
 PI Whitmore TE;  
 XX  
 DR WPI; 1999-095324/08.  
 XX  
 PT New mammalian Zneu1 polypeptides - used to, e.g. treat Alzheimer's  
 PT disease, cancer and to repopulate blood cells



XX Claim 6; Page 48-49; 70pp; English.

PS  
XX  
CC This polypeptide comprises human zneul mature polypeptide. Zneul  
CC is a new neuro-growth factor-like protein (see also W88381). Its  
CC closest human homologue is HSMHCW5A, a gene in the HLA class III  
CC region, which is contained in a cosmid which contains Notch 4.  
CC Zneul is also homologous to Notch 4 in its EGF-like domains and may  
CC be involved in EGF receptor pathways. Zneul is widely expressed in  
CC adult tissues, with high expression in heart, placenta, spleen,  
CC testis, thymoid, spinal cord and lymph node. Zneul can be used as  
CC a growth, maintenance, or differentiation factor in the spinal  
CC cord, heart, spleen, testis, thymoid and lymph nodes. It may  
CC play a role in breast cancer, glioblastomas, and pituitary  
CC adenomas. Zneul may also be used to treat Alzheimer's disease,  
CC cancer, to repopulate blood cells after chemotherapy, to stimulate  
CC myofibroblast proliferation, stimulate or inhibit growth factors  
CC made in the placenta, in fertility and contraception, or to  
CC regenerate nerves. Claimed zneul polypeptides (see also W88382-97),  
CC including specific domains of zneul and epitope-bearing portions of  
CC zneul, can be used to raise specific antibodies for use e.g. in  
CC diagnostic assays.

XX  
SQ Sequence 254 AA;

Query Match 87.2%; Score 41; DB 20; Length 254;  
Best Local Similarity 87.5%; Pred. No. 17;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GGLCVOPG 8  
Db 95 ggcscvpg 102  
111111

RESULT 12  
B42204  
ID B42204 standard; Protein; 265 AA.

XX  
AC B42204;

XX  
DT 08-FEB-2001 (first entry)

XX  
DE Human ORFX ORF1968 polypeptide sequence SEQ ID NO:3936.

XX  
KW Human: open reading frame; ORFX: detection; cytostatic; hepatotropic;  
KW vulnery; antipsoriatic; antiparkinsonian; neurotropic; neuroprotective;  
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiac;  
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;  
KW hypotensive; dermatological; immunosuppressive; antinflammatory;  
KW antiviral; antibacterial; antifungal; antineumatic; antithyroid;  
KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;  
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;  
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;  
KW cholesterol ester storage; systemic lupus erythematosus; infection;  
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;  
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;  
KW bone damage; cartilage damage; antinflammatory disease; coagulation;  
KW thrombosis; contraceptive.

XX  
OS Homo sapiens.

XX  
PN WO200058473-A2.

XX  
PD 05-OCT-2000.

XX  
PE 31-MAR-2000; 2000WO-US08621.

XX  
PR 31-MAR-1999; 99US-0127607.  
PR 02-APR-1999; 99US-0127636.  
PR 05-APR-1999; 99US-0127728.  
PR 30-MAR-2000; 2000US-0540763.

PA (CURA-) CURAGEN CORP.

XX  
XX Shimkels RA, Leach M;  
PI  
XX  
DR WPI: 2000-602362/57.  
XX  
XX N-PSDS; C76413.

PT Novel nucleic acids and peptides derived from open reading frame X,  
PT useful for treating e.g. cancers, proliferative disorders,  
PT neurodegenerative disorders and cardiovascular disease -

XX  
PS Claim 11; Page 3102; 5507pp; English.

XX  
CC C74446 to C77606 encode the proteins given in B40237 to B43397, which  
CC represent the human ORFX open reading frames 1 to 3161. The ORFX  
CC sequences have activities such as: cytostatic; hepatotropic; vulnery;  
CC antipsoriatic; antiparkinsonian; neurotropic; neuroprotective; osteopathic;  
CC anticonvulsant; antiarthritic; immunosuppressant; immunostimulant;  
CC cardiant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive;  
CC dermatological; immunosuppressive; antinflammatory; antibacterial;  
CC antiviral; antifungal; antineumatic; antithyroid; and antianaemic. The  
CC sequences can be used for determining the presence of or predisposition  
CC to, or preventing or treating pathological conditions associated with an  
CC ORFX-associated disorder. The nucleic acids can be used to express ORFX  
CC proteins in gene therapy vectors. The proteins and nucleic acids may be  
CC used to treat cancers, proliferative disorders, neurodegenerative  
CC disorders, osteoarthritis, graft vs host disease, cardiovascular disease,  
CC diabetes mellitus, hypertension, hypothyroidism, cholesterol ester  
CC storage, systemic lupus erythematosus, severe combined immunodeficiency  
CC (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune  
CC disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and  
CC cartilage damage, nocturnal haemoglobinuria, antinflammatory disease; to  
CC enhance coagulation; to inhibit thrombosis; and as a contraceptive.

XX  
SQ Sequence 265 AA;

Query Match 87.2%; Score 41; DB 21; Length 265;  
Best Local Similarity 87.5%; Pred. No. 18;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GGLCVOPG 8  
Db 114 ggcscvpg 121  
111111

RESULT 13  
Y41769  
ID Y41769 standard; Protein; 273 AA.

XX  
AC Y41769;

XX  
DT 07-DEC-1999 (first entry)

XX  
DE Human PRO13-1 protein sequence.

XX  
KW Human: PRO; EST; expressed sequence tag; PCR primer; hybridisation;  
KW probe; blood coagulation disorder; cancer; cellular adhesion disorder;  
KW secreted protein; transmembrane protein.

XX  
OS Homo sapiens.

XX  
PN WO9946281-A2.

XX  
PD 16-SEP-1999.

XX  
PE 08-MAR-1999; 99WO-US05028.

XX  
PR 10-MAR-1998; 98US-0077450.  
PR 11-MAR-1998; 98US-0077632.  
PR 11-MAR-1998; 98US-0077641.  
PR 11-MAR-1998; 98US-0077649.  
PR 12-MAR-1998; 98US-0077791.

PR 13-MAR-1998; 98US-0078004.  
 PR 17-MAR-1998; 98US-0040220.  
 PR 20-MAR-1998; 98US-0078886.  
 PR 20-MAR-1998; 98US-0078910.  
 PR 20-MAR-1998; 98US-0078936.  
 PR 20-MAR-1998; 98US-0078939.  
 PR 25-MAR-1998; 98US-0079294.  
 PR 26-MAR-1998; 98US-0079656.  
 PR 27-MAR-1998; 98US-0079663.  
 PR 27-MAR-1998; 98US-0079664.  
 PR 27-MAR-1998; 98US-0079689.  
 PR 27-MAR-1998; 98US-0079728.  
 PR 27-MAR-1998; 98US-0079786.  
 PR 30-MAR-1998; 98US-0079920.  
 PR 30-MAR-1998; 98US-0079923.  
 PR 31-MAR-1998; 98US-0080105.  
 PR 31-MAR-1998; 98US-0080107.  
 PR 31-MAR-1998; 98US-0080165.  
 PR 31-MAR-1998; 98US-0080194.  
 PR 01-APR-1998; 98US-0080327.  
 PR 01-APR-1998; 98US-0080328.  
 PR 01-APR-1998; 98US-0080333.  
 PR 01-APR-1998; 98US-0080334.  
 PR 08-APR-1998; 98US-0081049.  
 PR 08-APR-1998; 98US-0081070.  
 PR 08-APR-1998; 98US-0081071.  
 PR 09-APR-1998; 98US-0081195.  
 PR 09-APR-1998; 98US-0081203.  
 PR 09-APR-1998; 98US-0081229.  
 PR 15-APR-1998; 98US-0081817.  
 PR 15-APR-1998; 98US-0081838.  
 PR 15-APR-1998; 98US-0081952.  
 PR 15-APR-1998; 98US-0081955.  
 PR 21-APR-1998; 98US-0082568.  
 PR 21-APR-1998; 98US-0082569.  
 PR 22-APR-1998; 98US-0082700.  
 PR 22-APR-1998; 98US-0082704.  
 PR 22-APR-1998; 98US-0082804.  
 PR 23-APR-1998; 98US-0082767.  
 PR 23-APR-1998; 98US-0082796.  
 PR 27-APR-1998; 98US-0083336.  
 PR 28-APR-1998; 98US-0083322.  
 PR 29-APR-1998; 98US-0083392.  
 PR 29-APR-1998; 98US-0083495.  
 PR 29-APR-1998; 98US-0083496.  
 PR 29-APR-1998; 98US-0083499.  
 PR 29-APR-1998; 98US-0083500.  
 PR 29-APR-1998; 98US-0083545.  
 PR 29-APR-1998; 98US-0083554.  
 PR 29-APR-1998; 98US-0083558.  
 PR 29-APR-1998; 98US-0083559.  
 PR 30-APR-1998; 98US-0083742.  
 PR 05-MAY-1998; 98US-0084366.  
 PR 06-MAY-1998; 98US-0084414.  
 PR 06-MAY-1998; 98US-0084411.  
 PR 07-MAY-1998; 98US-0084598.  
 PR 07-MAY-1998; 98US-0084600.  
 PR 07-MAY-1998; 98US-0084627.  
 PR 07-MAY-1998; 98US-0084637.  
 PR 07-MAY-1998; 98US-0084639.  
 PR 07-MAY-1998; 98US-0084640.  
 PR 07-MAY-1998; 98US-0084643.  
 PR 13-MAY-1998; 98US-0085323.  
 PR 13-MAY-1998; 98US-0085338.  
 PR 13-MAY-1998; 98US-0085339.  
 PR 15-MAY-1998; 98US-0085573.  
 PR 15-MAY-1998; 98US-0085579.  
 PR 15-MAY-1998; 98US-0085580.  
 PR 15-MAY-1998; 98US-0085582.  
 PR 15-MAY-1998; 98US-0085689.  
 PR 15-MAY-1998; 98US-0085697.  
 PR 15-MAY-1998; 98US-0085700.  
 PR 15-MAY-1998; 98US-0085704.

PR 18-MAY-1998; 98US-0086023.  
 PR 22-MAY-1998; 98US-0086392.  
 PR 22-MAY-1998; 98US-0086414.  
 PR 22-MAY-1998; 98US-0086430.  
 PR 22-MAY-1998; 98US-0086436.  
 PR 28-MAY-1998; 98US-0087098.  
 PR 28-MAY-1998; 98US-0087106.  
 PR 28-MAY-1998; 98US-0087208.  
 PR 30-JUL-1998; 98US-0094651.  
 PR 11-SEP-1998; 98US-0100038.  
 PA (GETH ) GENENTECH INC;  
 PI Wood WI, Goddard A, Gurney A, Yuan J, Baker KP, Chen J;  
 XX WPI. 1999-551358/46.  
 DR N-PSDB: 234311.  
 PT New secreted and transmembrane polypeptides and their polynucleotides,  
 PT useful for treating blood coagulation disorders, cancers and cellular  
 PT adhesion disorders -  
 XX  
 PS Claim 12; Fig 213; 530pp; English.  
 CC The present invention describes secreted and transmembrane polypeptides  
 CC and their polynucleotides. The nucleotide sequences are useful as  
 CC sources of probes, primers, for chromosome mapping, and for generation  
 CC of antisense sequences. They can also be used to create transgenic  
 CC animals. The proteins can be used to treat a variety of diseases and  
 CC disorders, depending on their function. Diseases that may be treated  
 CC include blood coagulation disorders, cancers and cellular adhesion  
 CC disorders. They may also be used to raise antibodies. 233891 to  
 CC 234338, and Y41774 represent polynucleotide and polypeptide  
 CC sequence given in the exemplification of the present invention.  
 XX  
 SQ Sequence 273 AA;  
 Query Match 87.2%; Score 41; DB 20; Length 273;  
 Best Local Similarity 87.5%; Pred. No. 18;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 GGLCVOPG 8  
 II IIIII  
 Db 114 ggcvcvpg 121  
 RESULT 14  
 Y41770  
 ID Y41770 standard; Protein: 273 AA.  
 XX  
 AC Y41770;  
 XX  
 DT 07-DEC-1999 (first entry)  
 XX  
 DE Human PRO1330 protein sequence.  
 XX  
 KM Human; PRO; EST; expressed sequence tag; PCR primer; hybridisation;  
 KW probe; blood coagulation disorder; cancer; cellular adhesion disorder;  
 XX secreted protein; transmembrane protein.  
 OS Homo sapiens.  
 XX  
 PN W09946281-A2.  
 XX  
 PD 16-SEP-1999.  
 XX  
 PF 08-MAR-1999; 99WO-US05028.  
 XX  
 PR 10-MAR-1998; 98US-0077450.  
 PR 11-MAR-1998; 98US-0077632.  
 PR 11-MAR-1998; 98US-0077641.  
 PR 11-MAR-1998; 98US-0077649.

```

PR 12-MAR-1998; 98US-0077791.
PR 13-MAR-1998; 98US-0078004.
PR 17-MAR-1998; 98US-0040220.
PR 20-MAR-1998; 98US-0078886.
PR 20-MAR-1998; 98US-0078910.
PR 20-MAR-1998; 98US-0078936.
PR 20-MAR-1998; 98US-0078939.
PR 25-MAR-1998; 98US-0079294.
PR 26-MAR-1998; 98US-0079636.
PR 27-MAR-1998; 98US-0079663.
PR 27-MAR-1998; 98US-0079664.
PR 27-MAR-1998; 98US-0079689.
PR 27-MAR-1998; 98US-0079728.
PR 27-MAR-1998; 98US-0079786.
PR 30-MAR-1998; 98US-0079920.
PR 30-MAR-1998; 98US-0079923.
PR 31-MAR-1998; 98US-0080105.
PR 31-MAR-1998; 98US-0080107.
PR 31-MAR-1998; 98US-0080165.
PR 31-MAR-1998; 98US-0080194.
PR 01-APR-1998; 98US-0080327.
PR 01-APR-1998; 98US-0080328.
PR 01-APR-1998; 98US-0080333.
PR 01-APR-1998; 98US-0080334.
PR 08-APR-1998; 98US-0081049.
PR 08-APR-1998; 98US-0081070.
PR 08-APR-1998; 98US-0081071.
PR 09-APR-1998; 98US-0081195.
PR 09-APR-1998; 98US-0081203.
PR 09-APR-1998; 98US-0081229.
PR 15-APR-1998; 98US-0081817.
PR 15-APR-1998; 98US-0081838.
PR 15-APR-1998; 98US-0081955.
PR 15-APR-1998; 98US-0081956.
PR 21-APR-1998; 98US-0082568.
PR 22-APR-1998; 98US-0082700.
PR 22-APR-1998; 98US-0082704.
PR 22-APR-1998; 98US-0082804.
PR 23-APR-1998; 98US-0082767.
PR 23-APR-1998; 98US-0082796.
PR 27-APR-1998; 98US-0083336.
PR 28-APR-1998; 98US-0083322.
PR 29-APR-1998; 98US-0083392.
PR 29-APR-1998; 98US-0083495.
PR 29-APR-1998; 98US-0083496.
PR 29-APR-1998; 98US-0083499.
PR 29-APR-1998; 98US-0083500.
PR 29-APR-1998; 98US-0083545.
PR 29-APR-1998; 98US-0083554.
PR 29-APR-1998; 98US-0083558.
PR 29-APR-1998; 98US-0083559.
PR 30-APR-1998; 98US-0083742.
PR 05-MAY-1998; 98US-0083742.
PR 06-MAY-1998; 98US-0084366.
PR 06-MAY-1998; 98US-0084414.
PR 06-MAY-1998; 98US-0084441.
PR 07-MAY-1998; 98US-0084598.
PR 07-MAY-1998; 98US-0084600.
PR 07-MAY-1998; 98US-0084627.
PR 07-MAY-1998; 98US-0084637.
PR 07-MAY-1998; 98US-0084639.
PR 07-MAY-1998; 98US-0084640.
PR 07-MAY-1998; 98US-0084643.
PR 13-MAY-1998; 98US-0085323.
PR 13-MAY-1998; 98US-0085338.
PR 13-MAY-1998; 98US-0085339.
PR 15-MAY-1998; 98US-0085573.
PR 15-MAY-1998; 98US-0085579.
PR 15-MAY-1998; 98US-0085580.
PR 15-MAY-1998; 98US-0085582.
PR 15-MAY-1998; 98US-0085689.
PR 15-MAY-1998; 98US-0085697.
PR 15-MAY-1998; 98US-0085700.

PR 15-MAY-1998; 98US-0077791.
PR 13-MAR-1998; 98US-0078004.
PR 17-MAR-1998; 98US-0040220.
PR 20-MAR-1998; 98US-0078886.
PR 20-MAR-1998; 98US-0078910.
PR 20-MAR-1998; 98US-0078936.
PR 20-MAR-1998; 98US-0078939.
PR 25-MAR-1998; 98US-0079294.
PR 26-MAR-1998; 98US-0079636.
PR 27-MAR-1998; 98US-0079663.
PR 27-MAR-1998; 98US-0079664.
PR 27-MAR-1998; 98US-0079689.
PR 27-MAR-1998; 98US-0079728.
PR 27-MAR-1998; 98US-0079786.
PR 30-MAR-1998; 98US-0079920.
PR 30-MAR-1998; 98US-0079923.
PR 31-MAR-1998; 98US-0080105.
PR 31-MAR-1998; 98US-0080107.
PR 31-MAR-1998; 98US-0080165.
PR 31-MAR-1998; 98US-0080194.
PR 01-APR-1998; 98US-0080327.
PR 01-APR-1998; 98US-0080328.
PR 01-APR-1998; 98US-0080333.
PR 01-APR-1998; 98US-0080334.
PR 08-APR-1998; 98US-0081049.
PR 08-APR-1998; 98US-0081070.
PR 08-APR-1998; 98US-0081071.
PR 09-APR-1998; 98US-0081195.
PR 09-APR-1998; 98US-0081203.
PR 09-APR-1998; 98US-0081229.
PR 15-APR-1998; 98US-0081817.
PR 15-APR-1998; 98US-0081838.
PR 15-APR-1998; 98US-0081955.
PR 15-APR-1998; 98US-0081956.
PR 21-APR-1998; 98US-0082568.
PR 22-APR-1998; 98US-0082700.
PR 22-APR-1998; 98US-0082704.
PR 22-APR-1998; 98US-0082804.
PR 23-APR-1998; 98US-0082767.
PR 23-APR-1998; 98US-0082796.
PR 27-APR-1998; 98US-0083336.
PR 28-APR-1998; 98US-0083322.
PR 29-APR-1998; 98US-0083392.
PR 29-APR-1998; 98US-0083495.
PR 29-APR-1998; 98US-0083496.
PR 29-APR-1998; 98US-0083499.
PR 29-APR-1998; 98US-0083500.
PR 29-APR-1998; 98US-0083545.
PR 29-APR-1998; 98US-0083554.
PR 29-APR-1998; 98US-0083558.
PR 29-APR-1998; 98US-0083559.
PR 30-APR-1998; 98US-0083742.
PR 05-MAY-1998; 98US-0083742.
PR 06-MAY-1998; 98US-0084366.
PR 06-MAY-1998; 98US-0084414.
PR 06-MAY-1998; 98US-0084441.
PR 07-MAY-1998; 98US-0084598.
PR 07-MAY-1998; 98US-0084600.
PR 07-MAY-1998; 98US-0084627.
PR 07-MAY-1998; 98US-0084637.
PR 07-MAY-1998; 98US-0084639.
PR 07-MAY-1998; 98US-0084640.
PR 07-MAY-1998; 98US-0084643.
PR 13-MAY-1998; 98US-0085323.
PR 13-MAY-1998; 98US-0085338.
PR 13-MAY-1998; 98US-0085339.
PR 15-MAY-1998; 98US-0085573.
PR 15-MAY-1998; 98US-0085579.
PR 15-MAY-1998; 98US-0085580.
PR 15-MAY-1998; 98US-0085582.
PR 15-MAY-1998; 98US-0085689.
PR 15-MAY-1998; 98US-0085697.
PR 15-MAY-1998; 98US-0085700.

PR 15-MAY-1998; 98US-0085704.
PR 18-MAY-1998; 98US-0086023.
PR 22-MAY-1998; 98US-0086392.
PR 22-MAY-1998; 98US-0086414.
PR 22-MAY-1998; 98US-0086430.
PR 22-MAY-1998; 98US-0086439.
PR 28-MAY-1998; 98US-0087098.
PR 28-MAY-1998; 98US-0087106.
PR 28-MAY-1998; 98US-0087208.
PR 30-JUL-1998; 98US-0094651.
PR 11-SEP-1998; 98US-0100038.

PR 15-MAY-1998; 98US-0085704.
PR 18-MAY-1998; 98US-0086023.
PR 22-MAY-1998; 98US-0086392.
PR 22-MAY-1998; 98US-0086414.
PR 22-MAY-1998; 98US-0086430.
PR 22-MAY-1998; 98US-0086439.
PR 28-MAY-1998; 98US-0087098.
PR 28-MAY-1998; 98US-0087106.
PR 28-MAY-1998; 98US-0087208.
PR 30-JUL-1998; 98US-0094651.
PR 11-SEP-1998; 98US-0100038.

PR (GENE ) GENENTECH INC.
PR Wood WI, Goddard A, Gurney A, Yuan J, Baker KP, Chen J;
PR WPI; 1999-551358/46.
PR DR N-PSDB; 234312.
PR XX
PR PT New secreted and transmembrane polypeptides and their polynucleotides,
PR useful for treating blood coagulation disorders, cancers and cellular
PR adhesion disorders.
PR XX
PR PS Claim 12; Fig 215; 530pp; English.
PR XX
PR CC The present invention describes secreted and transmembrane polypeptides
PR and their polynucleotides. The nucleotide sequences are useful as
PR CC sources of probes, primers, for chromosome mapping, and for generation
PR CC of antisense sequences. They can also be used to create transgenic
PR CC animals. The proteins can be used to treat a variety of diseases and
PR CC disorders, depending on their function. Diseases that may be treated
PR CC include blood coagulation disorders, cancers and cellular adhesion
PR CC disorders. They may also be used to raise antibodies. Z33891 to
PR CC Z3438, and Y41685 to Y41774 represent polynucleotide and polypeptide
PR CC sequence given in the exemplification of the present invention.
PR XX
PR SQ Sequence 273 AA:

Query Match 87.2%; Score 41; DB 20; Length 273;
Best Local Similarity 87.5%; Pred. No. 18;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGLCVQPG 8
Db 114 gscvqpg 121

RESULT 15
Y41771
ID Y41771 standard; Protein; 273 AA.
XX
AC Y41771;
XX
DT 07-DEC-1999 (first entry)
XX
DE Human PRO1449 protein sequence.
XX
KW Human; PRO; EST; expressed sequence tag; PCR primer; hybridisation;
KW probe; blood coagulation disorder; cancer; cellular adhesion disorder;
KW secreted protein; transmembrane protein.
XX
OS Homo sapiens.
XX
PN WO9946281-A2.
XX
PD 16-SEP-1999.
XX
PF 08-MAR-1999; 99WO-US05028.
XX
PR 10-MAR-1998; 98US-0077450.
PR 11-MAR-1998; 98US-0077632.
PR 11-MAR-1998; 98US-0077641.

```

```

PR 11-MAR-1998; 98US-0077649.
PR 12-MAR-1998; 98US-0077791.
PR 13-MAR-1998; 98US-0078004.
PR 17-MAR-1998; 98US-0040222.
PR 20-MAR-1998; 98US-0078888.
PR 20-MAR-1998; 98US-0078910.
PR 20-MAR-1998; 98US-0078936.
PR 25-MAR-1998; 98US-0078939.
PR 25-MAR-1998; 98US-0079294.
PR 26-MAR-1998; 98US-0079656.
PR 27-MAR-1998; 98US-0079663.
PR 27-MAR-1998; 98US-0079668.
PR 27-MAR-1998; 98US-0079689.
PR 27-MAR-1998; 98US-0079728.
PR 27-MAR-1998; 98US-0079786.
PR 30-MAR-1998; 98US-0079920.
PR 30-MAR-1998; 98US-0079923.
PR 31-MAR-1998; 98US-0080105.
PR 31-MAR-1998; 98US-0080107.
PR 31-MAR-1998; 98US-0080165.
PR 31-MAR-1998; 98US-0080194.
PR 01-APR-1998; 98US-0080327.
PR 01-APR-1998; 98US-0080328.
PR 01-APR-1998; 98US-0080333.
PR 01-APR-1998; 98US-0080334.
PR 06-APR-1998; 98US-0081049.
PR 08-APR-1998; 98US-0081070.
PR 08-APR-1998; 98US-0081071.
PR 09-APR-1998; 98US-0081195.
PR 09-APR-1998; 98US-0081203.
PR 09-APR-1998; 98US-0081229.
PR 15-APR-1998; 98US-0081817.
PR 15-APR-1998; 98US-0081838.
PR 15-APR-1998; 98US-0081952.
PR 15-APR-1998; 98US-0081955.
PR 21-APR-1998; 98US-0082568.
PR 21-APR-1998; 98US-0082569.
PR 22-APR-1998; 98US-0082700.
PR 22-APR-1998; 98US-0082704.
PR 22-APR-1998; 98US-0082804.
PR 23-APR-1998; 98US-0082767.
PR 23-APR-1998; 98US-0082796.
PR 27-APR-1998; 98US-0083336.
PR 28-APR-1998; 98US-0083332.
PR 29-APR-1998; 98US-0083392.
PR 29-APR-1998; 98US-0083495.
PR 29-APR-1998; 98US-0083496.
PR 29-APR-1998; 98US-0083499.
PR 29-APR-1998; 98US-0083500.
PR 29-APR-1998; 98US-0083545.
PR 29-APR-1998; 98US-0083554.
PR 29-APR-1998; 98US-0083558.
PR 29-APR-1998; 98US-0083559.
PR 30-APR-1998; 98US-0083742.
PR 05-MAY-1998; 98US-0084366.
PR 06-MAY-1998; 98US-0084414.
PR 06-MAY-1998; 98US-0084441.
PR 07-MAY-1998; 98US-0084598.
PR 07-MAY-1998; 98US-0084600.
PR 07-MAY-1998; 98US-0084627.
PR 07-MAY-1998; 98US-0084637.
PR 07-MAY-1998; 98US-0084639.
PR 07-MAY-1998; 98US-0084640.
PR 07-MAY-1998; 98US-0084643.
PR 13-MAY-1998; 98US-0085323.
PR 13-MAY-1998; 98US-0085338.
PR 13-MAY-1998; 98US-0085339.
PR 15-MAY-1998; 98US-0085573.
PR 15-MAY-1998; 98US-0085579.
PR 15-MAY-1998; 98US-0085580.
PR 15-MAY-1998; 98US-0085582.
PR 15-MAY-1998; 98US-0085689.
PR 15-MAY-1998; 98US-0085697.

PR 15-MAY-1998; 98US-0085700.
PR 15-MAY-1998; 98US-0085704.
PR 18-MAY-1998; 98US-0086023.
PR 22-MAY-1998; 98US-0086392.
PR 22-MAY-1998; 98US-0086414.
PR 22-MAY-1998; 98US-0086430.
PR 22-MAY-1998; 98US-0086486.
PR 28-MAY-1998; 98US-0087098.
PR 28-MAY-1998; 98US-0087106.
PR 28-MAY-1998; 98US-0087208.
PR 30-JUL-1998; 98US-0094651.
PR 11-SEP-1998; 98US-0100038.

PA (GETH ) GENENTECH INC.
PI Wood WI, Goddard A, Gurney A, Yuan J, Baker KP, Chen J;
XX WPI; 1999-551358/46.
XX N-PSDB; 234313.
XX
PT New secreted and transmembrane polypeptides and their polynucleotides,
PT useful for treating blood coagulation disorders, cancers and cellular
PT adhesion disorders .
XX
XX Claim 12; Fig 217; 530pp; English.
XX
XX The present invention describes secreted and transmembrane polypeptides
XX and their polynucleotides. The nucleotide sequences are useful as
XX sources of probes, primers, for chromosome mapping, and for generation
XX of antisense sequences. They can also be used to create transgenic
XX animals. The proteins can be used to treat a variety of diseases and
XX disorders, depending on their function. Diseases that may be treated
XX include blood coagulation disorders, cancers and cellular adhesion
XX disorders. They may also be used to raise antibodies. Z33891 to
XX Z34338, and Y41685 to Y41774 represent polynucleotide and polypeptide
XX sequence given in the exemplification of the present invention.
XX
XX Sequence 273 AA:

Query Match 87.2%; Score 41; DB 20; Length 273;
Best Local Similarity 87.5%; Pred. No. 18;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GGLCVQPG 8
Db 114 gsgcvqpg 121

Search completed: June 13, 2001, 14:14:39
Job time: 380 sec

```



*This Page Blank (uspto)*

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 13, 2001, 14:10:46 ; Search time 130.61 Seconds  
(without alignments)  
4.209 Million cell updates/sec

Title: PCT-US01-05825A-17

Perfect score: 47  
Sequence: 1 GGLCVQPG 8

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 198801 seqs, 68722935 residues

Total number of hits satisfying chosen parameters: 198801

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_67:\*

1: p1r1:.\*  
2: p1r2:.\*  
3: p1r3:.\*  
4: p1r4:.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	41	87.2	379	2 A59180	Wnt inhibitory fac
2	39	83.0	378	2 B59180	Wnt inhibitory fac
3	38	80.9	278	2 E65020	ethanolamine u111
4	37	78.7	250	2 T10512	hypothetical prote
5	37	78.7	423	2 S39830	hypothetical prote
6	37	78.7	427	2 D83347	probable aminotran
7	35	74.5	241	2 S26615	nuclear factor I-A
8	35	74.5	300	1 B70627	hypothetical prote
9	35	74.5	309	2 E36596	nuclear factor I (
10	35	74.5	347	2 E36596	nuclear factor I (
11	35	74.5	357	2 PC4293	nuclear factor I f
12	35	74.5	373	2 A70856	probable lpp2 prot
13	35	74.5	386	2 S45569	nuclear factor I-A
14	35	74.5	431	2 S20065	nuclear factor I-A
15	35	74.5	433	2 S51773	transcription fact
16	35	74.5	441	2 S50112	nuclear factor I-X
17	35	74.5	441	2 B31256	transcription fact
18	35	74.5	461	2 S45568	nuclear factor I-A
19	35	74.5	498	2 S45567	nuclear factor I-A
20	35	74.5	509	2 S45565	nuclear factor I-A
21	35	74.5	509	2 JC5428	nuclear factor I-A
22	35	74.5	522	2 S09996	nuclear factor I-A
23	35	74.5	531	2 JN0586	beta-1,4-mannosyl-
24	35	74.5	532	2 B36596	nuclear factor I -
25	35	74.5	536	2 A43415	beta-1,4-mannosyl-
26	35	74.5	536	2 UC4362	beta-1,4-mannosyl-
27	35	74.5	780	2 T03156	ribonucleoside-dip
28	35	74.5	1184	2 G01763	atrophin-1 - human
29	35	74.5	1184	2 S50832	atrophin-1 - human

30	34	72.3	98	2 S26929	Ig heavy chain V r
31	34	72.3	120	2 S44111	Ig heavy chain V-D
32	34	72.3	196	2 F71525	hypothetical prote
33	34	72.3	196	2 H81681	mat protein TC0628
34	34	72.3	302	2 T44578	C-5 ketoreductase
35	34	72.3	331	2 T28374	ORF MSV213 hypotbe
36	34	72.3	403	2 E69873	cell-division prot
37	34	72.3	421	2 A38168	glutamate dehydrog
38	34	72.3	467	2 D72461	hypothetical prote
39	34	72.3	505	2 H83196	glycerol kinase PA
40	34	72.3	861	2 A48825	Notch homolog Molc
41	34	72.3	948	2 E82206	DnaK-related prote
42	34	72.3	2531	2 A46019	Notch-1 protein -
43	34	72.3	2555	2 A40043	notch protein homo
44	33	70.2	145	2 H69051	heterodisulfide re
45	33	70.2	324	2 S61447	cellulase (EC 3.2.

ALIGNMENTS

RESULT 1  
A59180  
Wnt inhibitory factor-1 - human  
C:Species: Homo sapiens (man)  
C:Date: 18-Feb-2000 #sequence\_revision 18-Feb-2000 #text\_change 18-Feb-2000  
R:Accession: A59180  
R:Hisleh, J.C.; Kodjabachian, L.; Rebber, M.L.; Ratner, A.; Smallwood, P.M.; Samos, Nature 398, 431-436, 1999  
A:Title: A new secreted protein that binds to Wnt proteins and inhibits their activit  
A:Reference number: A59180; MUID:99215557  
A:Accession: A59180  
A:Status: preliminary; not compared with conceptual translation  
A:Molecule type: mRNA  
A:Residues: 1-379 <HS1>  
A:Cross-references: GB:AF122922; NID:g4585369; PIDN:AAD25402.1; PID:g4585370

Query Match 87.2%; Score 41; DB 2; Length 379;  
Best Local Similarity 87.5%; Pred. No. 2.9;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGLCVQPG 8  
DB 221 GGLCVTPG 228  
RESULT 2  
B59180  
Wnt inhibitory factor-1 - zebra fish  
C:Species: Brachydanio rerio (zebra fish)  
C:Date: 18-Feb-2000 #sequence\_revision 18-Feb-2000 #text\_change 18-Feb-2000  
R:Accession: B59180  
R:Hisleh, J.C.; Kodjabachian, L.; Rebber, M.L.; Ratner, A.; Smallwood, P.M.; Samos, Nature 398, 431-436, 1999  
A:Title: A new secreted protein that binds to Wnt proteins and inhibits their activit  
A:Reference number: A59180; MUID:99215557  
A:Accession: B59180  
A:Status: preliminary; not compared with conceptual translation  
A:Molecule type: mRNA  
A:Residues: 1-378 <HS1>  
A:Cross-references: GB:AF122925; NID:g4585375; PIDN:AAD25405.1; PID:g4585376  
A:Gene: WIF-1  
Query Match 83.0%; Score 39; DB 2; Length 378;  
Best Local Similarity 75.0%; Pred. No. 7;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
QY 1 GGLCVQPG 8  
DB 219 GGLCVSPG 226

RESULT 3  
E65020  
ethanolamine utilization protein EutY - Escherichia coli (strain K-12)  
C:Species: Escherichia coli  
C:Date: 12-Sep-1997 #sequence\_revision 17-Sep-1997 #text\_change 08-Oct-1999  
C:Accession: E65020  
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; CD  
A.; Rose, D.J.; Mau, B.; Shao, Y.  
Science 277, 1453-1462, 1997  
A:Title: The complete genome sequence of Escherichia coli K-12.  
A:Reference number: A64720; MUID:97426617  
A:Accession: E65020  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-278 <BLAT>  
A:Cross-references: GB:AE000332; GB:U00096; NID:q1788789; PIDN:AACT5507.1; PID:q1788796;  
A:Experimental source: strain K-12, substrain MG1655  
C:Genetics:  
A:Gene: eutY

Query Match 80.9%; Score 38; DB 2; Length 278;  
Best Local Similarity 75.0%; Pred. No. 8.1;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGLCVOPG 8  
|||:||||  
DB 229 GGSCMQPG 236

RESULT 4  
T10512  
hypothetical protein - Rhodobacter capsulatus  
C:Species: Rhodobacter capsulatus  
C:Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 16-Jul-1999  
C:Accession: T10512  
R:LaRimer, F.W.  
submitted to the EMBL Data Library, November 1995  
A:Description: Sequence and expression of the pentose-5'-phosphate 3'-phosphatase (cbbE) cat  
A:Reference number: Z17063  
A:Accession: T10512  
A:Status: preliminary; translated from GB/EMBL/DBD  
A:Molecule type: DNA  
A:Residues: 1-250 <LAR>  
A:Cross-references: EMBL:U23145; NID:g2564972; PID:g2564978  
A:Experimental source: strain ATCC 11166

Query Match 78.7%; Score 37; DB 2; Length 250;  
Best Local Similarity 85.7%; Pred. No. 11;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GGLCVOPG 8  
|||:||||  
DB 138 GLCAQPG 144

RESULT 5  
S39830  
hypothetical protein YBL058w - yeast (Saccharomyces cerevisiae)  
N:Alternate names: hypothetical protein YBL0509; hypothetical protein YBL0515  
C:Species: Saccharomyces cerevisiae  
C:Date: 16-Dec-1993 #sequence\_revision 09-Sep-1994 #text\_change 29-Oct-1999  
C:Accession: S39830; S45793; S37331  
R:Schereus, B.; el Bakoury, M.; Vierendeels, F.; Dubois, E.; Messenguy, F.  
Yeast 9, 1355-1371, 1993  
A:Title: Sequencing and functional analysis of a 32 560 bp segment on the left arm of ye  
A:Reference number: S39824; MUID:94205266  
A:Accession: S39830  
A:Molecule type: DNA  
A:Residues: 1-423 <SCH>

A:Cross-references: EMBL:Z23261; NID:g313733; PIDN:CAA80789.1; PID:g313740  
A:Experimental source: strain S286C  
R:Dubois, E.; El Bakoury, M.; Glansdorff, N.; Messenguy, F.; Pierard, A.; Schereus,  
submitted to the Protein Sequence Database, August 1994  
A:Reference number: S45782  
A:Accession: S45793  
A:Molecule type: DNA  
A:Residues: 1-423 <DUB>

A:Cross-references: EMBL:Z235819; NID:g536089; PIDN:CAA84878.1; PID:g536090; MIPS:YBL0  
A:Experimental source: strain S286C  
C:Genetics:  
A:Gene: SCD:SHPI  
A:Cross-references: SCD:S0000154; MIPS:YBL058w  
A:Map position: 2L

Query Match 78.7%; Score 37; DB 2; Length 423;  
Best Local Similarity 75.0%; Pred. No. 18;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGLCVOPG 8  
|||:||||  
DB 85 GGLCPKPG 92

RESULT 6  
D83347  
probable aminotransferase PA2394 [imported] - Pseudomonas aeruginosa (strain PA01)  
C:Species: Pseudomonas aeruginosa  
C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
C:Accession: D83347  
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.;  
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L  
.; Lory, S.; Olson, M.V.  
Nature 406, 959-964, 2000  
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa  
A:Reference number: AB2950; MUID:20437337  
A:Accession: D83347  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-427 <STO>  
A:Cross-references: GB:AE004665; GB:AE004091; NID:g9948426; PIDN:AA05782.1; GSPDB:GN  
A:Experimental source: strain PA01  
C:Genetics:  
A:Gene: PA2394

Query Match 78.7%; Score 37; DB 2; Length 427;  
Best Local Similarity 87.5%; Pred. No. 19;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGLCVOPG 8  
|||:||||  
DB 131 GGLVOPG 138

RESULT 7  
S26615  
nuclear factor I-A3 - chicken  
C:Species: Gallus gallus (chicken)  
C:Date: 03-May-1994 #sequence\_revision 27-Jan-1995 #text\_change 08-Dec-2000  
C:Accession: S26615; S45566  
R:Kruse, U.  
submitted to the EMBL Data Library, August 1992  
A:Reference number: S26614  
A:Accession: S26615  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-241 <KRU>  
A:Cross-references: EMBL:X68153; NID:g63665; PID:g63666  
J:Kruse, U.; Sippel, A.E.  
J. Mol. Biol. 238, 860-865, 1994  
A:Title: The genes for transcription factor nuclear factor I give rise to correspondi



A:Reference number: S45565; MUID:94238700  
A:Accession: S45566  
A:Status: preliminary; nucleic acid sequence not shown  
A:Molecule type: mRNA  
A:Residues: 2-202 <KR2>  
A:Cross-references: EMBL:X68153  
C:Superfamily: human nuclear factor 1

Query Match 74.5%; Score 35; DB 2; Length 241;  
Best Local Similarity 100.0%; Pred. No. 26;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLCVOP 7  
|||||  
DB 152 GLCVOP 157

RESULT 8  
B70627  
hypothetical protein RV0263c - Mycobacterium tuberculosis (strain H37RV)  
C:Species: Mycobacterium tuberculosis  
C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 21-Jul-2000  
C:Accession: B70627  
R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Felkell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998  
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
A:Reference number: A70500; MUID:98295987  
A:Accession: B70627  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-300 <COL>  
A:Cross-references: GB:286089; GB:AL123456; NID:g3261711; PIDN:CAB06690.1; PID:g1850108  
A:Experimental source: strain H37RV  
C:Genetics:  
A:Gene: RV0263c  
C:Superfamily: hypothetical protein H11730

Query Match 74.5%; Score 35; DB 1; Length 300;  
Best Local Similarity 71.4%; Pred. No. 32;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 GLCVOP 7  
|||:  
DB 120 GCVCVP 126

RESULT 9  
E36596  
nuclear factor 1 (clone B5) - mouse (fragment)  
C:Species: Mus musculus (house mouse)  
C:Date: 12-Apr-1991 #sequence\_revision 12-Apr-1991 #text\_change 08-Dec-2000  
C:Accession: E36596  
R:Inoue, T.; Tamura, T.; Furutachi, T.; Mikoshiba, K.; J. Biol. Chem. 265, 19065-19070, 1990  
A:Title: Isolation of complementary DNAs encoding a cerebellum-enriched nuclear factor 1  
A:Reference number: A36596; MUID:91035434  
A:Accession: E36596  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-309 <INO>  
A:Cross-references: GB:D90172; GB:M58637  
C:Superfamily: human nuclear factor 1

Query Match 74.5%; Score 35; DB 2; Length 309;  
Best Local Similarity 100.0%; Pred. No. 33;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLCVOP 7  
|||||  
DB 69 GLCVOP 74

RESULT 10  
F36596  
nuclear factor 1 (clone B6) - mouse  
C:Species: Mus musculus (house mouse)  
C:Date: 12-Apr-1991 #sequence\_revision 12-Apr-1991 #text\_change 08-Dec-2000  
C:Accession: F36596  
R:Inoue, T.; Tamura, T.; Furutachi, T.; Mikoshiba, K.; J. Biol. Chem. 265, 19065-19070, 1990  
A:Title: Isolation of complementary DNAs encoding a cerebellum-enriched nuclear factor 1  
A:Reference number: A36596; MUID:91035434  
A:Accession: F36596  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-347 <INO>  
A:Cross-references: GB:D90176; GB:M58641  
C:Superfamily: human nuclear factor 1

Query Match 74.5%; Score 35; DB 2; Length 347;  
Best Local Similarity 100.0%; Pred. No. 37;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLCVOP 7  
|||||  
DB 160 GLCVOP 165

RESULT 11  
PC4293  
nuclear factor 1 family protein A2 - rat (fragment)  
C:Species: Rattus norvegicus (Norway rat)  
C:Date: 07-Jul-1997 #sequence\_revision 17-Sep-1997 #text\_change 08-Dec-2000  
C:Accession: PC4293  
R:Osada, S.; Daimon, S.; Ikeda, T.; Nishihara, T.; Yano, K.; Yamasaki, M.; Imagawa, M.; J. Biochem. 121, 355-363, 1997  
A:Title: Nuclear factor 1 family proteins bind to the silencer element in the rat glu  
A:Reference number: JCS428; MUID:97244636  
A:Accession: PC4293  
A:Molecule type: mRNA  
A:Residues: 1-357 <OSA>  
A:Cross-references: DBJ:D78020; NID:g1041037; PIDN:BA11206.1; PID:g1041038  
A:Experimental source: liver  
C:Comment: This protein is a dominant form of nuclear factor 1 proteins. It occupies  
ferase P gene expression.  
C:Superfamily: human nuclear factor 1

Query Match 74.5%; Score 35; DB 2; Length 357;  
Best Local Similarity 100.0%; Pred. No. 38;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLCVOP 7  
|||||  
DB 19 GLCVOP 24

RESULT 12  
A70856  
probable lpp2 protein - Mycobacterium tuberculosis (strain H37RV)  
C:Species: Mycobacterium tuberculosis  
C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 22-Oct-1999  
C:Accession: A70856  
R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Felkell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998  
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno

A:Reference number: A70500; MUID:98295987  
A:Accession: A70856  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-373 <COL>  
A:Cross-references: GB:AL021287; GB:AL123456; NID:93261508; PIDN:CAA16091.1; PID:e123775  
A:Experimental source: strain H37Rv  
C:Genetics:  
A:Gene: lpp2

Query Match 74.5%; Score 35; DB 2; Length 373;  
Best Local Similarity 71.4%; Pred. No. 39;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 GGLCVOP 7  
|||||  
Db 234 GGLCIDP 240

RESULT 13  
S45569  
nuclear factor I-A6 - chicken  
C:Species: Gallus gallus (chicken)  
C>Date: 28-Oct-1994 #sequence\_revision 28-Oct-1994 #text\_change 08-Dec-2000  
C:Accession: S45569; S26618  
R:Krusse, U.; Stippel, A.E.  
J. Mol. Biol. 238, 860-865, 1994  
A:Title: The genes for transcription factor nuclear factor I give rise to corresponding  
A:Reference number: S45565; MUID:94238700  
A:Accession: S45569  
A:Status: preliminary; nucleic acid sequence not shown  
A:Molecule type: mRNA  
A:Residues: 1-386 <KKU>  
A:Cross-references: EMBL:X68156; NID:963671; PIDN:CAA48259.1; PID:963672  
C:Superfamily: human nuclear factor I

Query Match 74.5%; Score 35; DB 2; Length 386;  
Best Local Similarity 100.0%; Pred. No. 40;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GGLCVOP 7  
|||||  
Db 148 GGLCVOP 153

RESULT 14  
S20065  
nuclear factor I-X - chicken  
N:Alternate names: transcription factor X  
C:Species: Gallus gallus (chicken)  
C>Date: 22-Nov-1993 #sequence\_revision 10-Nov-1995 #text\_change 08-Dec-2000  
C:Accession: S20065  
R:Krusse, U.; Qian, F.; Stippel, A.E.  
Nucleic Acids Res. 19, 6641, 1991  
A:Title: Identification of a fourth nuclear factor I gene in chicken by cDNA cloning: NF  
A:Reference number: S20065; MUID:92093631  
A:Accession: S20065  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: mRNA  
A:Residues: 1-431 <KKU>  
A:Cross-references: EMBL:X61225; NID:963683; PIDN:CAA43537.1; PID:963684  
A:Note: this sequence was submitted to the EMBL Data Library, August 1991  
C:Superfamily: human nuclear factor I

Query Match 74.5%; Score 35; DB 2; Length 431;  
Best Local Similarity 100.0%; Pred. No. 45;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GGLCVOP 7  
|||||

Db 160 GGLCVOP 165

RESULT 15  
S51773  
transcription factor NF1-X2 - African clawed frog  
C:Species: Xenopus laevis (African clawed frog)  
C>Date: 07-May-1995 #sequence\_revision 03-Nov-1995 #text\_change 08-Dec-2000  
C:Accession: S51773  
R:roulet, E.; Armentero, M.; Dreyer, C.; Mermod, N.; Wahli, W.  
submitted to the EMBL Data Library, June 1994  
A:Description: DNA binding activity of xenopus nuclear factor X is negatively regulat  
A:Reference number: S51773  
A:Accession: S51773  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-433 <KOU>  
A:Cross-references: EMBL:Z34463; NID:9607204; PID:9607205  
C:Superfamily: human nuclear factor I

Query Match 74.5%; Score 35; DB 2; Length 433;  
Best Local Similarity 100.0%; Pred. No. 45;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GGLCVOP 7  
|||||  
Db 152 GGLCVOP 157

Search completed: June 13, 2001, 14:10:47  
job time: 148 sec



*This Page Blank (uspto)*

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 13, 2001, 14:21:47 ; Search time 74.44 Seconds  
(without alignments)  
3.681 Million cell updates/sec

Title: PCT-US01-05825A-17  
Perfect score: 47  
Sequence: 1 GGLCVQPG 8

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues  
Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_39:\*

Prod. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	41	87.2	379	1	WIF1_HUMAN
2	41	87.2	379	1	WIF1_MOUSE
3	38	80.9	278	1	EUTJ_ECOLI
4	38	80.9	279	1	EUTJ_SALTY
5	37	78.7	423	1	SHP1_YEAST
6	35	74.5	431	1	NFX1_CHICK
7	35	74.5	441	1	NFX1_HUMAN
8	35	74.5	441	1	NFX1_MOUSE
9	35	74.5	441	1	NFX1_MOUSE
10	35	74.5	509	1	NFX1_MOUSE
11	35	74.5	522	1	NFX1_MOUSE
12	35	74.5	531	1	NFX1_MOUSE
13	35	74.5	536	1	NFX1_MOUSE
14	35	74.5	536	1	NFX1_MOUSE
15	35	74.5	536	1	NFX1_MOUSE
16	35	74.5	536	1	NFX1_MOUSE
17	35	74.5	536	1	NFX1_MOUSE
18	35	74.5	536	1	NFX1_MOUSE
19	35	74.5	536	1	NFX1_MOUSE
20	35	74.5	536	1	NFX1_MOUSE
21	35	74.5	536	1	NFX1_MOUSE
22	35	74.5	536	1	NFX1_MOUSE
23	35	74.5	536	1	NFX1_MOUSE
24	35	74.5	536	1	NFX1_MOUSE
25	35	74.5	536	1	NFX1_MOUSE
26	35	74.5	536	1	NFX1_MOUSE
27	35	74.5	536	1	NFX1_MOUSE
28	35	74.5	536	1	NFX1_MOUSE
29	35	74.5	536	1	NFX1_MOUSE
30	35	74.5	536	1	NFX1_MOUSE
31	35	74.5	536	1	NFX1_MOUSE
32	35	74.5	536	1	NFX1_MOUSE
33	35	74.5	536	1	NFX1_MOUSE

## ALIGNMENTS

RESULT	ID	STANDARD	PRT	379 AA
1	WIF1_HUMAN			
AC	Q9Y5W5			
DT	01-OCT-2000 (Rel. 40, Created)			
DR	01-OCT-2000 (Rel. 40, Last sequence update)			
DT	01-OCT-2000 (Rel. 40, Last annotation update)			
DE	WNT INHIBITORY FACTOR 1 PRECURSOR (WIF-1).			
GN	WIF1.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.			
OX	NCBI_Taxid=9606;			
RP	SEQUENCE FROM N.A.			
RN	MEDLINE=9215557; PubMed=10201374;			
RX	Smith J.-C., Kodjabachian L., Rebert M.L., Ratner A.,			
RA	Smallwood P.M., Samos C.H., Nusse R., David I.B., Nathans J.;			
RT	"A new secreted protein that binds to Wnt proteins and inhibits their			
FT	activities."			
RL	Nature 398:431-436(1999).			
CC	- FUNCTION: BINDS TO WNT PROTEINS AND INHIBITS THEIR ACTIVITIES. MAY			
CC	BE INVOLVED IN MESODERM SEGMENTATION.			
CC	- SUBCELLULAR LOCATION: SECRETED.			
CC	- SIMILARITY: CONTAINS 5 EGF-LIKE DOMAINS.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See <a href="http://www.isb-sdb.ch/announce/">http://www.isb-sdb.ch/announce/</a>			
CC	or send an email to <a href="mailto:license@sdb.ch">license@sdb.ch</a> ).			
CC	EMBL: AF122922; AAD25402.1; -			
DR	MIM: 605186; -			
DR	HSSP: P00743; ICFP.			
DR	InterPro: IPR000561; -			
DR	InterPro: IPR002049; -			
DR	Pfam: PF00008; EGF_5.			
DR	PRINTS: PR00011; EGFAMININ.			
DR	PROSITE: PS00022; EGF_1; 5.			
DR	PROSITE: PS01186; EGF_2; 4.			
KW	Repeat: EGF-like domain; Signal: Developmental protein.			
FT	SIGNAL	1	28	POTENTIAL.
FT	CHAIN	29	379	WNT INHIBITORY FACTOR 1.
FT	DOMAIN	177	208	EGF-LIKE 1.
FT	DOMAIN	209	240	EGF-LIKE 2.
FT	DOMAIN	241	272	EGF-LIKE 3.
FT	DOMAIN	273	304	EGF-LIKE 4.
FT	DOMAIN	305	336	EGF-LIKE 5.
FT	DISULFID	177	186	POTENTIAL.
FT	DISULFID	182	192	POTENTIAL.
FT	DISULFID	198	200	POTENTIAL.
FT	DISULFID	209	218	POTENTIAL.
FT	DISULFID	214	224	POTENTIAL.

FT DISULFID 230 232 POTENTIAL.  
 FT DISULFID 241 250 POTENTIAL.  
 FT DISULFID 246 256 POTENTIAL.  
 FT DISULFID 262 264 POTENTIAL.  
 FT DISULFID 273 282 POTENTIAL.  
 FT DISULFID 278 288 POTENTIAL.  
 FT DISULFID 294 296 POTENTIAL.  
 FT DISULFID 305 314 POTENTIAL.  
 FT DISULFID 310 320 POTENTIAL.  
 FT DISULFID 326 328 POTENTIAL.  
 FT CARBOHYD 88 88 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 245 245 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 379 AA: 41512 MW: 27782370A266E784 CRC64;

Query Match 87.2%; Score 41; DB 1; Length 379;  
 Best Local Similarity 87.5%; Pred. No. 1.8;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GGLCVQPG 8  
 |||||  
 |||||  
 DB 221 GGLCVTPG 228

RESULT 2  
 ID WIFL\_MOUSE STANDARD; PRT: 379 AA.  
 AC Q9W0AL;  
 DT 01-OCT-2000 (Rel. 40, Created)  
 DT 01-OCT-2000 (Rel. 40, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE WNT INHIBITORY FACTOR 1 PRECURSOR (WIF-1).  
 GN WIF1.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP MEDLINE=99215557; PubMed=10201374;  
 RA Hasleh J.-C., Kodjabachian L., Rebber M.L., Rattner A.,  
 RA Smallwood P.M., Samos C.H., Nusse R., David I.B., Nathans J.;  
 RT "A new secreted protein that binds to Wnt proteins and inhibits their  
 activities.";  
 RL Nature 398:431-436(1999).  
 CC -1- FUNCTION: BINDS TO WNT PROTEINS AND INHIBITS THEIR ACTIVITIES. MAY  
 BE INVOLVED IN MESODERM SEGMENTATION.  
 CC -1- SUBCELLULAR LOCATION: SECRETED.  
 CC -1- TISSUE SPECIFICITY: EXPRESSION HIGHEST IN HEART AND LUNG. LOWER IN  
 BRAIN AND EYE.  
 CC -1- SIMILARITY: CONTAINS 5 EGF-LIKE DOMAINS.  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: AF122923; AAD25403.1; -  
 CC HSSP: P00740; 11XA. -  
 CC MGD: MG1:1344332; Wifl. -  
 CC InterPro: IPR000561; -  
 CC InterPro: IPR002049; -  
 CC Pfam: PF00008; EGF\_5. -  
 CC PRINTS: PR00011; EGF\_LAMININ. -  
 CC PROSITE: PS00022; EGF\_1; 5. -  
 CC PROSITE: PS01186; EGF\_2; 4. -  
 CC Repeat: EGF-like domain; Signal: Developmental protein.  
 FT SIGMUL 1 28  
 FT CHAIN 29 379  
 FT DOMAIN 177 208 WNT INHIBITORY FACTOR 1.  
 CC EGF-LIKE 1.

FT DOMAIN 209 240 EGF-LIKE 2.  
 FT DOMAIN 241 272 EGF-LIKE 3.  
 FT DOMAIN 273 304 EGF-LIKE 4.  
 FT DOMAIN 305 336 EGF-LIKE 5.  
 FT DISULFID 177 186 POTENTIAL.  
 FT DISULFID 182 192 POTENTIAL.  
 FT DISULFID 198 200 POTENTIAL.  
 FT DISULFID 209 218 POTENTIAL.  
 FT DISULFID 214 224 POTENTIAL.  
 FT DISULFID 230 232 POTENTIAL.  
 FT DISULFID 241 250 POTENTIAL.  
 FT DISULFID 246 256 POTENTIAL.  
 FT DISULFID 262 264 POTENTIAL.  
 FT DISULFID 273 282 POTENTIAL.  
 FT DISULFID 278 288 POTENTIAL.  
 FT DISULFID 294 296 POTENTIAL.  
 FT DISULFID 305 314 POTENTIAL.  
 FT DISULFID 310 320 POTENTIAL.  
 FT DISULFID 326 328 POTENTIAL.  
 FT CARBOHYD 88 88 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 245 245 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 379 AA: 41590 MW: E3765F2642B2BC9A CRC64;

Query Match 87.2%; Score 41; DB 1; Length 379;  
 Best Local Similarity 87.5%; Pred. No. 1.8;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GGLCVQPG 8  
 |||||  
 |||||  
 DB 221 GGLCVTPG 228

RESULT 3  
 ID EUTJ\_ECOLI STANDARD; PRT: 278 AA.  
 AC P72727;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE ETHANOLAMINE UTILIZATION PROTEIN EUTJ.  
 GN EUTJ.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Escherichia.  
 OX NCBI\_TaxID=562;  
 RN [1]  
 RP MEDLINE=97426617; PubMed=9278503;  
 RC STRAIN=K12 / MG1655;  
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
 RA Mau B., Shao Y.;  
 RT "The complete genome sequence of Escherichia coli K-12.";  
 RL Science 277:1453-1474(1997).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12;  
 RA Alba H., Baba T., Fujita K., Hayashi K., Honjo A., Horiiuchi T.,  
 RA Ikemoto K., Inada T., Isono K., Itoh T., Kanai K., Kasai H.,  
 RA Kashimoto K., Kim S., Kimura S., Kitagawa M., Kitakawa M., Makino K.,  
 RA Masuda S., Miki T., Mizobuchi K., Mori H., Motomura K., Nakamura Y.,  
 RA Nishimoto H., Nishio Y., Oshima T., Saito N., Sampei G., Seki Y.,  
 RA Tagami H., Takemoto K., Wada C., Yamamoto Y., Yano M.;  
 RL Submitted (JAN-1997) to the EMBL/Genbank/DBJ databases.  
 CC -1- PATHWAY: ETHANOLAMINE UTILIZATION.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial



OC Gallus.  
 ON NCBI\_TaxID=9031;  
 RX [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92093631; PubMed=1754401;  
 RA Kruse U., Qian F., Sippel A.E.;  
 RT "Identification of a fourth nuclear factor I gene in chicken by cDNA cloning: NF-I-X".  
 RL Nucleic Acids Res. 19:6641-6641(1991).  
 CC -1- FUNCTION: CTF/NF-I RECOGNIZE AND BIND THE PALINDROMIC SEQUENCE 5'-TTGGCANNNGGCA-3' PRESENT IN VIRAL AND CELLULAR PROMOTERS AND IN THE ORIGIN OF REPLICATION OF ADENOVIRUS TYPE 2. THESE PROTEINS ARE INDIVIDUALLY CAPABLE OF ACTIVATING TRANSCRIPTION AND REPLICATION.  
 CC -1- SUBUNIT: BINDS DNA AS AN HOMODIMER.  
 CC -1- ALTERNATIVE PRODUCTS: DIFFERENT FORMS OF CTF/NF-I CAN BE PRODUCED BY DIFFERENTIAL SPLICING OF THE SAME GENE.  
 CC -1- SIMILARITY: BELONGS TO THE CTF/NF-I FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X61225; CAA43537.1; -  
 DR InterPro: IPR000647; -  
 DR Pfam: PF00859; CTF\_NFI; 1.  
 DR PROSITE: PS00349; CTF\_NFI; 1.  
 KW Transcription regulation; DNA replication; DNA-binding; Activator;  
 KW Nuclear protein; Multigene family; Alternative splicing.  
 SQ SEQUENCE 431 AA; 46903 MW; 83989A0292E74D6C CRC64;  
 QY Query Match 74.5%; Score 35; DB 1; Length 431;  
 Best Local Similarity 100.0%; Pred. No. 25;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 2 GLCVP 7  
 Db 160 GLCVP 165  
 RESULT 7  
 NFIX\_HUMAN  
 ID NFIX\_HUMAN STANDARD; PRT; 441 AA.  
 AC 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE NUCLEAR FACTOR I/X (NF-I/X) (CCAAT-BOX BINDING TRANSCRIPTION FACTOR) (CTF) (TGCGA-BINDING PROTEIN).  
 GN NFIX.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OC NCBI\_TaxID=9606;  
 OX [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=foreskin;  
 RX MEDLINE=95023136; PubMed=7937100;  
 RA Apt D., Liu Y., Bernard H.U.;  
 RT "Cloning and functional analysis of spliced isoforms of human nuclear factor I-X: interference with transcriptional activation by NF1/CTF in a cell-type specific manner".  
 RL Nucleic Acids Res. 22:3825-3833(1994).  
 RN [2]  
 RP SEQUENCE OF 19-228 FROM N.A.  
 RX MEDLINE=96070434; PubMed=7590749;  
 RA Qian F., Kruse U., Lichter P., Sippel A.E.;  
 RT "Chromosomal localization of the four genes (NFIA, B, C, and X) for

RT the human transcription factor nuclear factor I by FISH".  
 RL Genomics 28:66-73(1995).  
 CC -1- FUNCTION: CTF/NF-I RECOGNIZE AND BIND THE PALINDROMIC SEQUENCE 5'-TTGGCANNNGGCA-3' PRESENT IN VIRAL AND CELLULAR PROMOTERS AND IN THE ORIGIN OF REPLICATION OF ADENOVIRUS TYPE 2. THESE PROTEINS ARE INDIVIDUALLY CAPABLE OF ACTIVATING TRANSCRIPTION AND REPLICATION.  
 CC -1- SUBUNIT: BINDS DNA AS AN HOMODIMER.  
 CC -1- ALTERNATIVE PRODUCTS: DIFFERENT FORMS OF CTF/NF-I CAN BE PRODUCED BY DIFFERENTIAL SPLICING OF THE SAME GENE.  
 CC -1- SIMILARITY: BELONGS TO THE CTF/NF-I FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; L31881; AAA3442.1; -  
 DR EMBL; U07811; AAA3126.1; -  
 DR MIM; 164005; -  
 DR InterPro: IPR000647; -  
 DR Pfam: PF00859; CTF\_NFI; 1.  
 DR PROSITE: PS00349; CTF\_NFI; 1.  
 KW Transcription regulation; DNA replication; DNA-binding; Activator;  
 KW Nuclear protein; Multigene family; Alternative splicing.  
 FT VARSPPLIC 318 359  
 FT CONFLICT 91 91  
 FT MISSING (IN ISOFORM NF1-X2).  
 SQ SEQUENCE 441 AA; 48883 MW; 304AF8993B9668E CRC64;  
 QY Query Match 74.5%; Score 35; DB 1; Length 441;  
 Best Local Similarity 100.0%; Pred. No. 26;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 2 GLCVP 7  
 Db 160 GLCVP 165  
 RESULT 8  
 NFIX\_MESAU  
 ID NFIX\_MESAU STANDARD; PRT; 441 AA.  
 AC P13623;  
 DT 01-JAN-1990 (Rel. 13, Created)  
 DT 01-JAN-1990 (Rel. 13, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE NUCLEAR FACTOR I/X (NF-I/X) (CCAAT-BOX BINDING TRANSCRIPTION FACTOR) (CTF) (TGCGA-BINDING PROTEIN).  
 GN NFIX.  
 OS Mesocricetus auratus (Golden hamster).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae; Mesocricetus.  
 OC NCBI\_TaxID=10036;  
 OX [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=89057859; PubMed=3194401;  
 RA Gil G., Smith J.R., Goldstein J.L., Slaughter C.A., Orth K., Brown M.S., Osborne T.F.;  
 RT "Multiple genes encode nuclear factor I-like proteins that bind to the promoter for 3-hydroxy-3-methylglutaryl-coenzyme A reductase".  
 RL Proc. Natl. Acad. Sci. U.S.A. 85:8963-8967(1988).  
 CC -1- FUNCTION: CTF/NF-I RECOGNIZE AND BIND THE PALINDROMIC SEQUENCE 5'-TTGGCANNNGGCA-3' PRESENT IN VIRAL AND CELLULAR PROMOTERS AND IN THE ORIGIN OF REPLICATION OF ADENOVIRUS TYPE 2. THESE PROTEINS ARE INDIVIDUALLY CAPABLE OF ACTIVATING TRANSCRIPTION AND REPLICATION.  
 CC -1- SUBUNIT: BINDS DNA AS AN HOMODIMER.  
 CC -1- ALTERNATIVE PRODUCTS: DIFFERENT FORMS OF CTF/NF-I CAN BE PRODUCED BY DIFFERENTIAL SPLICING OF THE SAME GENE.  
 CC -----



CC -1- SIMILARITY: BELONGS TO THE CTF/NF-I FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: J04123; AAA37083.1; -  
CC PIR: B31256; B31256.  
CC TRANSFAC: T00610; -  
CC TRANSFAC: T01035; -  
CC InterPro: IPR000647; -  
CC Pfam: PF00859; CTF\_NFI; 1.  
CC PROSITE: PS00349; CTF\_NFI; 1.  
CC Transcription regulation: DNA replication; DNA-binding; Activator;  
CC Nuclear protein; Multigene family; Alternative splicing.  
CC SEQUENCE 441 AA: 48837 MW: 38CB7326EDC92859 CRC64;  
KW  
SO  
Query Match 74.5%; Score 35; DB 1; Length 441;  
Best Local Similarity 100.0%; Pred. No. 26;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 2 GLCVOP 7  
Db 160 GLCVOP 165  
IIIIII  
RESULT 9  
NFIL\_MOUSE STANDARD; PRT; 441 AA.  
ID NFIL\_MOUSE  
AC P70257; P70258; O08519;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE NUCLEAR FACTOR 1/X (NF-1/X) (CCAAT-BOX BINDING TRANSCRIPTION  
DE FACTOR) (CTF) (TGCCA-BINDING PROTEIN).  
GN NF1X.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_Taxid=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA STRAIN-NIH SWISS;  
RC Ebel T.T., Sippl A.E.;  
RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE OF 4-441 FROM N.A.  
RC STRAIN-BALB/C;  
RX MEDLINE=97209336; PubMed=9056636;  
RA Chaudhry A.Z., Lyons G.E., Gronostajski R.M.;  
RT "Expression patterns of the four nuclear factor I genes during mouse  
RT embryogenesis indicate a potential role in development.";  
RL Dev. Dyn. 208:313-325(1997).  
CC -1- FUNCTION: CTF/NF-I RECOGNIZE AND BIND THE PALINDROMIC SEQUENCE 5'-  
CC TGGCANNNGCCAA-3' PRESENT IN VIRAL AND CELLULAR PROMOTERS AND IN  
CC THE ORIGIN OF REPLICATION OF ADENOVIRUS TYPE 2. THESE PROTEINS ARE  
CC INDIVIDUALLY CAPABLE OF ACTIVATING TRANSCRIPTION AND REPLICATION.  
CC -1- SUBUNIT: BINDS DNA AS AN HOMODIMER.  
CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
CC -1- ALTERNATIVE PRODUCTS: DIFFERENT FORMS OF CTF/NF-I CAN BE PRODUCED  
CC BY DIFFERENTIAL SPLICING OF THE SAME GENE.  
CC -1- SIMILARITY: BELONGS TO THE CTF/NF-I FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: Y07688; CAA68952.1; -  
CC DR EMBL: Y07689; CAA68953.1; -  
CC DR EMBL: U57636; AAB49931.1; -  
CC MCD: MGI:97311; Nf1x.  
CC InterPro: IPR000647; -  
CC Pfam: PF00859; CTF\_NFI; 1.  
CC PROSITE: PS00349; CTF\_NFI; 1.  
CC Transcription regulation: DNA replication; DNA-binding; Activator;  
CC Nuclear protein; Multigene family; Alternative splicing.  
CC VARSPLIC 320 360 MISSING (IN ISOFORM NF1X2).  
CC SEQUENCE 441 AA: 48867 MW: F0FCB00C2CD52480 CRC64;  
KW  
SO  
Query Match 74.5%; Score 35; DB 1; Length 441;  
Best Local Similarity 100.0%; Pred. No. 26;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 2 GLCVOP 7  
Db 160 GLCVOP 165  
IIIIII  
RESULT 10  
NFIL\_RAT STANDARD; PRT; 509 AA.  
ID NFIL\_RAT  
AC P09414; Q63782;  
DT 01-MAR-1989 (Rel. 10, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE NUCLEAR FACTOR 1 (NF-1) (CCAAT BOX-BINDING TRANSCRIPTION FACTOR) (CTF)  
DE (TGCCA-BINDING PROTEIN) (NFI-A).  
GN NF1L21.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_Taxid=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-SPRAGUE-DAWLEY; TISSUE=Liver;  
RX MEDLINE=96429753; PubMed=8832903;  
RA Monaci P., Nuzzo M., Stampelli S., Tollervey D., de Simone V.,  
RA Nicosia A.;  
RT "A complex interplay of positive and negative elements is responsible  
RT for the different transcriptional activity of liver NF1 variants.";  
RL Mol. Biol. Rep. 21:147-158(1995).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-SPRAGUE-DAWLEY; TISSUE=Liver;  
RX MEDLINE=97244636; PubMed=9089412;  
RA Osada S., Daimon S., Ikeda T., Nishihara T., Yano K., Yamasaki M.,  
RA Imagawa M.;  
RT "Nuclear factor 1 family proteins bind to the silencer element in the  
RT rat glutathione transferase P gene.";  
RL J. Biochem. 121:355-363(1997).  
RN [3]  
RP SEQUENCE OF 5-509 FROM N.A., AND PARTIAL SEQUENCE.  
RC TISSUE=Liver;  
RX MEDLINE=89030607; PubMed=3053160;  
RA Paonessa G., Gounari F., Frank R., Cortese R.;  
RT "Purification of a NF1-like DNA-binding protein from rat liver and  
RT cloning of the corresponding cDNA.";  
RL EMBD J. 7:3115-3123(1988).  
CC -1- FUNCTION: CTF/NF-I RECOGNIZE AND BIND THE PALINDROMIC SEQUENCE 5'-  
CC TGGCANNNGCCAA-3' PRESENT IN VIRAL AND CELLULAR PROMOTERS AND IN  
CC THE ORIGIN OF REPLICATION OF ADENOVIRUS TYPE 2. THESE PROTEINS ARE  
CC INDIVIDUALLY CAPABLE OF ACTIVATING TRANSCRIPTION AND REPLICATION.  
CC -1- SUBUNIT: BINDS DNA AS AN HOMODIMER.  
CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
CC -1- ALTERNATIVE PRODUCTS: DIFFERENT FORMS OF CTF/NF-I CAN BE PRODUCED  
CC BY DIFFERENTIAL SPLICING OF THE SAME GENE.

CC -1- SIMILARITY: BELONGS TO THE CTF/NFI FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; X84209; CA58995.1; -  
DR EMBL; D78017; BA01203.1; -  
DR EMBL; X13167; CA03155.1; -  
DR PIR; S01300; S01300.  
DR TRANSFAC; T00599; -  
DR InterPro; IPR000647; -  
DR Pfam; PF00859; CTF\_NFI; 1.  
DR PROSITE; PS00349; CTF\_NFI; 1.  
DR Transcription regulation; DNA replication; DNA-binding; Activator;  
DR Nuclear protein; Multigene family; Alternative splicing.  
KW Nucleic acid; 509 AA; 55976 MW; 59120C7090229F87 CRC64;  
SQ SEQUENCE 509 AA; 55976 MW; 59120C7090229F87 CRC64;

Query Match 74.5%; Score 35; DB 1; Length 509;  
Best Local Similarity 100.0%; Pred. No. 29;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 2 GLCVP 7  
Db 160 GLCVP 165  
RESULT 11  
ID NFI-CHICK STANDARD; PRT; 522 AA.  
AC .PI7923;  
DT 01-NOV-1990 (Rel. 16, Created)  
DT 01-NOV-1990 (Rel. 16, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE NUCLEAR FACTOR 1 A1 (NFI-1) (CCAAT BOX-BINDING TRANSCRIPTION FACTOR)  
DE (CTF) (TGCGA-BINDING PROTEIN).  
GN NFI-A1.  
OS Gallus gallus (Chicken).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
OC Gallus.  
OX NCBI\_TaxID=9031;  
RN [1]  
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
RC TISSUE=Embryo;  
RX MEDLINE-90251434; PubMed-2339052;  
RA Rupp R.A.W., Kruse U., Muthaup G., Goebel U., Beyreuther K.,  
RA Stippel A.E.;  
RT "Chicken NFI/TGCGA proteins are encoded by at least three independent  
RT genes: NFI-A, NFI-B and NFI-C with homologues in mammalian genomes.";  
RL Nucleic Acids Res. 18:2607-2616(1990).  
CC -1- FUNCTION: CTF/NFI-1 RECOGNIZE AND BIND THE PALINDROMIC SEQUENCE 5'-  
CC TTGCGNNNNCCGA-3' PRESENT IN VIRAL AND CELLULAR PROMOTERS AND IN  
CC THE ORIGIN OF REPLICATION OF ADENOVIRUS TYPE 2. THESE PROTEINS ARE  
CC INDIVIDUALLY CAPABLE OF ACTIVATING TRANSCRIPTION AND REPLICATION.  
CC -1- SUBUNIT: BINDS DNA AS AN HOMODIMER.  
CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
CC -1- ALTERNATIVE PRODUCTS: DIFFERENT FORMS OF CTF/NFI-1 CAN BE PRODUCED  
CC BY DIFFERENTIAL SPLICING OF THE SAME GENE.  
CC -1- MISCELLANEOUS: IN CHICKEN NFI/TGCGA PROTEINS ARE ENCODED BY AT  
CC LEAST THREE INDEPENDENT GENES: NFI-A, NFI-B AND NFI-C.  
CC -1- SIMILARITY: BELONGS TO THE CTF/NFI-1 FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; X51486; CA035853.1; -  
DR PIR; S09996; S09996.  
DR TRANSFAC; T00544; -  
DR InterPro; IPR000647; -  
DR Pfam; PF00859; CTF\_NFI; 1.  
DR PROSITE; PS00349; CTF\_NFI; 1.  
DR Transcription regulation; DNA replication; DNA-binding; Activator;  
DR Nuclear protein; Multigene family; Alternative splicing.  
KW Nucleic acid; 522 AA; 57473 MW; C51F3E5FE6077740 CRC64;  
SQ SEQUENCE 522 AA; 57473 MW; C51F3E5FE6077740 CRC64;

Query Match 74.5%; Score 35; DB 1; Length 522;  
Best Local Similarity 100.0%; Pred. No. 30;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 2 GLCVP 7  
Db 160 GLCVP 165  
RESULT 12  
ID GNT3\_HUMAN STANDARD; PRT; 531 AA.  
AC .G09327;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 01-OCT-1996 (Rel. 34, Last annotation update)  
DE BETA-1,4-MANNOSYL-GLYCOPROTEIN BETA-1,4-N-ACETYLGLUCOSAMINYL-  
DE TRANSFERASE (EC 2.4.1.144) (N-GLYCOSYL-OLIGOSACCHARIDE-GLYCOPROTEIN  
DE N-ACETYLGLUCOSAMINYLTRANSFERASE III) (N-ACETYLGLUCOSAMINYLTRANSFERASE  
DE III) (GNT-III) (GLCNAc-T III).  
GN MGAT3 OR GNT3.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE-93380894; PubMed-8370666;  
RA Ihara Y., Nishikawa A., Toma T., Soejima H., Nishikawa N., Taniguchi N.;  
RT "DNA cloning, expression, and chromosomal localization of human N-  
RT acetylglucosaminyltransferase III (Gnt-III).";  
RL J. Biochem. 113:692-698(1993).  
CC -1- FUNCTION: IT IS INVOLVED IN THE REGULATION OF THE BIOSYNTHESIS AND  
CC BIOLOGICAL FUNCTION OF GLYCOPROTEIN OLIGOSACCHARIDES. CATALYZES  
CC THE ADDITION OF N-ACETYLGLUCOSAMINE IN BETA 1-4 LINKAGE TO THE  
CC BETA-LINKED MANNOSE OF THE TRIMANNOSE CORE OF N-LINKED SUGAR  
CC CHAINS. IT IS ONE OF THE MOST IMPORTANT ENZYMS INVOLVED IN THE  
CC REGULATION OF THE BIOSYNTHESIS OF GLYCOPROTEIN OLIGOSACCHARIDES.  
CC -1- CATALYTIC ACTIVITY: UDP-N-ACETYL-D-GLUCOSAMINE + N-ACETYL-BETA-D-  
CC GLUCOSAMINYL-1,2-ALPHA-D-MANNOSYL-1,3-(N-ACETYL-BETA-D-  
CC GLUCOSAMINYL-1,2-ALPHA-D-MANNOSYL-1,6)-BETA-D-MANNOSYL-1,4-N-  
CC ACETYL-BETA-D-GLUCOSAMINYL-R = UDP + N-ACETYL-BETA-D-GLUCOSAMINYL-  
CC 1,2-ALPHA-D-MANNOSYL-1,3-(N-ACETYL-BETA-D-GLUCOSAMINYL-1,2-ALPHA-  
CC D-MANNOSYL-1,6)-(N-ACETYL-BETA-D-GLUCOSAMINYL-1,4)-BETA-D-  
CC MANNOSE + N-ACETYL-BETA-D-GLUCOSAMINYL-R.  
CC -1- PATHWAY: GLYCOSYLATION.  
CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. GOLGI.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; D13789; BA02937.1; -  
DR PIR; JN0586; JN0586.  
MW; 604621; -

CC CHAINS. IT IS ONE OF THE MOST IMPORTANT ENZYMES INVOLVED IN THE  
CC REGULATION OF THE BIOSYNTHESIS OF GLYCOPROTEIN OLIGOSACCHARIDES.  
CC -1- CARPLYTIC ACTIVITY: UDP-N-ACETYL-D-GLUCOSAMINE + N-ACETYL-BETA-D-  
CC CAPLACTYCNX1 2,1-DUDP-D-MANNOSE-6-PHOSPHATE

```

CC -----
CC GLUCOSAMINYL-1,2-ALPHA-D-MANNOSYL-1,6)-BETA-D-MANNOSYL-1,4-N-
CC ACETYL-BETA-D-GLUCOSAMINYL-R = UDP + N-ACETYL-BETA-D-GLUCOSAMINYL-
CC 1,2-ALPHA-D-MANNOSYL-1,3-(N-ACETYL-BETA-D-GLUCOSAMINYL-1,2-ALPHA-
CC D-MANNOSYL-1,6)-(N-ACETYL-BETA-D-GLUCOSAMINYL-1,4)-BETA-D-
CC MANNOSYL-1,4-N-ACETYL-BETA-D-GLUCOSAMINYL-R.
CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. GOLGI.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: D10852; BAA01625.1; -.
CC PIR: A43415; A43415.
CC DR Transferase; Glycosyltransferase; Transmembrane; Signal-anchor;
CC Glycoprotein; Golgi stack.
CC KW DOMAIN 1 5 CYTOPLASMIC (POTENTIAL).
CC FT TRANSFER 1 21 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN).
CC FT DOMAIN 22 536 LUMENAL, CATALYTIC (POTENTIAL).
CC FT DOMAIN 33 84 PRO-RICH.
CC FT CARBOHYD 243 243 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 261 261 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 399 399 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC SEQUENCE 536 AA; 61762 MW; 81817BEB6C307438 CRC64;

Query Match 74.5%; Score 35; DB 1; Length 536;
Best Local Similarity 62.5%; Pred. No. 30;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGLCVPG 8
DB 112 GVCCKP 119

RESULT 15
DRPL_RAT ID DRPL_RAT STANDARD; PRT; 1183 AA.
AC P54258;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE ATROPHIN-1 (DENTATORUBRAL-PALLIDOLYSIAN ATROPHY PROTEIN).
GN DRPLA.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Cerebellum, and Striatum;
RX MEDLINE=97317138; PubMed=9173996;
RA Loev S.J., Margolis R.L., Young W.S., Li S.-H., Schilling G.,
RA Ashworth R.G., Ross C.A.;
RT "Cloning and expression of the rat atrophin-I (DRPLA disease gene)
RT homologue."
RL Neurobiol. Dis. 2:129-138(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain, Cerebellum, Hippocampus, and Substantia nigra;
RX MEDLINE=96081227; PubMed=8541849;
RA Schmitt I., Eppien J.T., Riess O.;
RT "Predominant neuronal expression of the gene responsible for
RT dentatorubral-pallidolysian atrophy (DRPLA) in rat."
RL Hum. Mol. Genet. 4:1619-1624(1995).
CC -1- TISSUE SPECIFICITY: PREDOMINANT NEURONAL EXPRESSION, ALTHOUGH
CC MARKEDLY REDUCED AMOUNTS ARE FOUND IN MOST OTHER TISSUES.
CC -1- DEVELOPMENTAL STAGE: SIMILAR EXPRESSION AT ALL DEVELOPMENT STAGES
CC (DAY 14.5 P.C., 17.5 P.C., NEWBORNS AND ADULTS).

```

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U31777; AAA80337.1; -.
CC DR EMBL: X89453; CAA61623.1; -.
CC DR InterPro: IPR002951; -.
CC DR PRINTS: PR01222; ATROPHIN.
CC FT DOMAIN 165 171 POLY-PRO.
CC FT DOMAIN 303 306 POLY-PRO.
CC FT DOMAIN 377 383 POLY-SER.
CC FT DOMAIN 387 391 POLY-SER.
CC FT DOMAIN 440 446 POLY-PRO.
CC FT DOMAIN 477 480 POLY-HIS.
CC FT DOMAIN 481 489 POLY-GLN.
CC FT DOMAIN 502 505 POLY-PRO.
CC FT DOMAIN 562 572 POLY-SER.
CC FT DOMAIN 702 705 POLY-PRO.
CC FT CONFLICT 455 455 N -> S (IN REF. 2).
CC FT CONFLICT 594 594 F -> L (IN REF. 2).
CC FT CONFLICT 689 689 P -> R (IN REF. 2).
CC FT CONFLICT 717 717 T -> M (IN REF. 2).
CC FT CONFLICT 737 737 A -> V (IN REF. 2).
CC FT CONFLICT 965 965 MISSING (IN REF. 2).
CC SEQUENCE 1183 AA; 124778 MW; 7FB9928DCAD9B1F CRC64;

Query Match 74.5%; Score 35; DB 1; Length 1183;
Best Local Similarity 75.0%; Pred. No. 60;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGLCVPG 8
DB 968 GGLALPG 975

```

Search completed: June 13, 2001, 14:21:48  
Job Time: 808 sec



This Page Blank (uspto)

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 13, 2001, 14:20:35 ; Search time 225.85 Seconds  
(without alignments)  
4.152 Million cell updates/sec

Title: PCT-US01-05825A-17  
Sequence: 1 GGLCVQPG 8

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 374700 seqs, 117207915 residues  
Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

SPREMBL\_15:\*

- 1: sp.archaea:\*
- 2: sp.bacteria:\*
- 3: sp.fungi:\*
- 4: sp.human:\*
- 5: sp.invertebrate:\*
- 6: sp.mammal:\*
- 7: sp.mhc:\*
- 8: sp.organelle:\*
- 9: sp.phage:\*
- 10: sp.plant:\*
- 11: sp.podent:\*
- 12: sp.unclassified:\*
- 13: sp.vertebrate:\*
- 14: sp.virus:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	41	87.2	273	4 Q9UHF1	Q9uhf1 homo sapien
2	41	87.2	374	13 Q9W6F8	Q9w6f8 xenopus lae
3	41	87.2	379	4 Q9T5W5	Q9t5w5 homo sapien
4	41	87.2	379	11 Q9WUAI	Q9wuai mus musculu
5	39	83.0	378	13 Q9W6F9	Q9w6f9 brachydanio
6	39	83.0	4123	4 Q75851	Q75851 mus musculu
7	38	80.9	368	11 Q9JLR0	Q9jlr0 mus musculu
8	38	80.9	381	4 Q755A3	Q755a3 homo sapien
9	38	80.9	474	4 Q9P0W3	Q9p0w3 homo sapien
10	38	80.9	487	4 Q9P0W5	Q9p0w5 homo sapien
11	38	80.9	870	6 Q02660	Q02660 bos taurus
12	37	78.7	250	2 Q33515	Q33515 rhodobacter
13	36	76.6	278	11 Q9OXT5	Q9oxt5 mus musculu
14	36	76.6	645	3 Q94106	Q94106 pneumocysti
15	36	76.6	647	3 Q94104	Q94104 pneumocysti
16	35	74.5	144	4 Q13051	Q13051 homo sapien
17	35	74.5	177	11 Q9R1G5	Q9r1g5 mus musculu
18	35	74.5	177	11 Q9R1G2	Q9r1g2 mus musculu
19	35	74.5	186	4 Q60413	Q60413 homo sapien

20	35	74.5	225	4 Q12857	Q12857 homo sapien
21	35	74.5	241	13 Q90926	Q90926 gallus gall
22	35	74.5	268	11 Q70190	Q70190 rattus norv
23	35	74.5	300	2 P95220	P95220 mycobacteri
24	35	74.5	357	11 Q63785	Q63785 rattus norv
25	35	74.5	373	2 Q53253	Q53253 mycobacteri
26	35	74.5	378	11 Q61960	Q61960 mus musculu
27	35	74.5	380	11 Q63784	Q63784 rattus norv
28	35	74.5	386	13 Q90929	Q90929 gallus gall
29	35	74.5	414	13 Q9PSA0	Q9psa0 xenopus lae
30	35	74.5	424	4 Q13052	Q13052 homo sapien
31	35	74.5	433	4 Q13050	Q13050 homo sapien
32	35	74.5	433	13 Q91522	Q91522 xenopus lae
33	35	74.5	461	13 Q90928	Q90928 gallus gall
34	35	74.5	466	11 P70251	P70251 mus musculu
35	35	74.5	487	11 Q63783	Q63783 rattus norv
36	35	74.5	488	11 Q64192	Q64192 mus sp. nfi
37	35	74.5	497	13 Q9PS99	Q9ps99 xenopus lae
38	35	74.5	498	11 Q63573	Q63573 rattus norv
39	35	74.5	498	13 Q90927	Q90927 gallus gall
40	35	74.5	509	11 P70250	P70250 mus musculu
41	35	74.5	509	13 Q90925	Q90925 gallus gall
42	35	74.5	531	4 Q9UH32	Q9uh32 homo sapien
43	35	74.5	532	11 Q02780	Q02780 mus musculu
44	35	74.5	532	11 Q90Y90	Q90y90 rattus norv
45	35	74.5	561	4 Q9P2A9	Q9p2a9 homo sapien

#### ALIGNMENTS

RESULT 1

Q9UHF1 ID Q9UHF1 PRELIMINARY: PRT: 273 AA.

AC Q9UHF1; ID 01-MAY-2000 (TREMBLrel. 13, Created)

DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)

DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)

DE NOTCH4-LIKE PROTEIN.

GN ZNEU1.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

OX NCBI\_Taxid=9606;

RN [1]

RA SHEPARD P., Jelinek L., Whitmore T., Blumberg H., Lehnert J.,

RA O'Hara P.

RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.

DR EMBL, AF186111; AA01429.1; ..

DR HSSP; P35555; IEMN.

DR INTERPRO; IPR000152; ..

DR INTERPRO; IPR000561; ..

DR INTERPRO; IPR001881; ..

DR PFAM; PF00008; EGF\_2.

DR PROSITE; PS00010; ASX\_HYDROXYL. 1.

DR PROSITE; PS00022; EGF\_1; UNKNOWN\_1.

DR PROSITE; PS01186; EGF\_2; UNKNOWN\_1.

DR PROSITE; PS01187; EGF\_CA; 1.

SO SEQUENCE 273 AA; 29617 MW; 5740BB845ED5A988 CRC64;

Query Match 87.2% Score 41; DB 4; Length 273;  
Best Local Similarity 87.5%; Pred. No. 3.2;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGLCVQPG 8  
DB 114 GGLCVQPG 121

RESULT 2

Q9W6F8

ID Q9W6F8 PRELIMINARY; PRT; 374 AA.  
 AC Q9W6F8;  
 DT 01-NOV-1999 (TRENBLREL. 12, Created)  
 DT 01-NOV-1999 (TRENBLREL. 12, Last sequence update)  
 DT 01-OCT-2000 (TRENBLREL. 15, Last annotation update)  
 DE WNT INHIBITORY FACTOR 1 PRECURSOR (WNT INHIBITORY FACTOR-1) (WIF-1).  
 GN WIF1 OR WIF-1.  
 OS Xenopus laevis (African clawed frog).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;  
 OC Xenopodinae; Xenopus.  
 OX NCBI\_TaxID=8355;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99215557; PubMed=10201374;  
 RA Hsieh J.-C., Kodjabachian L., Rebber M.L., Ratner A.,  
 RA Smallwood P.M., Samos C.H., Nusse R., David I.B., Nathans J.;  
 RT "A new secreted protein that binds to Wnt proteins and inhibits their  
 RT activities";  
 RL Nature 398:431-436(1999).  
 CC -1- FUNCTION: BINDS TO WNT PROTEINS AND INHIBITS THEIR ACTIVITIES. MAY  
 CC BE INVOLVED IN MESODERM SEGMENTATION.  
 CC -1- TISSUE SPECIFICITY: DURING SOMATOGENESIS, EXPRESSED PREDOMINANTLY  
 CC IN UNSEGMENTED PARAXIAL PRESOMITIC MESODERM AND, TO A MUCH LESSER  
 CC EXTENT, IN NEWLY SEGMENTED SOMITES.  
 CC -1- DEVELOPMENTAL STAGE: FIRST EXPRESSED AT NEURULA STAGES.  
 CC EMBL: AF122924; AAD25404.1; -.  
 DR HSSP: P00740; IEDM.  
 DR INTERPRO: IPR000561; -.  
 DR INTERPRO: IPR002049; -.  
 DR PRAM: PR00008; EGF. 5.  
 DR PRINTS: PR00011; EGF\_LAMININ.  
 DR PROSITE: PS00022; EGF\_1; UNKNOWN\_5.  
 DR PROSITE: PS01186; EGF\_2; 5.  
 KW Repeat; EGF-like domain; Signal; Developmental protein.  
 FT CHAIN 1 28  
 FT SIGNAL 1 28  
 FT DOMAIN 29 374  
 FT DOMAIN 172 203  
 FT DOMAIN 204 235  
 FT DOMAIN 236 267  
 FT DOMAIN 268 299  
 FT DOMAIN 300 331  
 FT DOMAIN 172 181  
 FT DISULFID 177 187  
 FT DISULFID 193 195  
 FT DISULFID 204 213  
 FT DISULFID 209 219  
 FT DISULFID 225 227  
 FT DISULFID 236 245  
 FT DISULFID 241 251  
 FT DISULFID 257 259  
 FT DISULFID 268 277  
 FT DISULFID 273 283  
 FT DISULFID 289 291  
 FT DISULFID 300 309  
 FT DISULFID 305 315  
 FT DISULFID 321 323  
 SQ SEQUENCE 374 AA; 41071 MW; E26F973B0F00ACF8 CRC64;

Query Match 87.2%; Score 41; DB 13; Length 374;  
 Best Local Similarity 87.5%; Pred. No. 4.2;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GGLCVOPG 8  
 DB 216 GGLCVTPG.223

RESULT 3  
 O9Y5W5 PRELIMINARY; PRT; 379 AA.

AC Q9Y5W5;  
 DT 01-NOV-1999 (TRENBLREL. 12, Created)  
 DT 01-NOV-1999 (TRENBLREL. 12, Last sequence update)  
 DT 01-OCT-2000 (TRENBLREL. 15, Last annotation update)  
 DE WNT INHIBITORY FACTOR 1 PRECURSOR (WNT INHIBITORY FACTOR-1) (WIF-1).  
 GN WIF1 OR WIF-1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99215557; PubMed=10201374;  
 RA Hsieh J.-C., Kodjabachian L., Rebber M.L., Ratner A.,  
 RA Smallwood P.M., Samos C.H., Nusse R., David I.B., Nathans J.;  
 RT "A new secreted protein that binds to Wnt proteins and inhibits their  
 RT activities";  
 RL Nature 398:431-436(1999).  
 CC -1- FUNCTION: BINDS TO WNT PROTEINS AND INHIBITS THEIR ACTIVITIES. MAY  
 CC BE INVOLVED IN MESODERM SEGMENTATION.  
 CC -1- SIMILARITY: CONTAINS 5 EGF-LIKE DOMAINS.  
 CC EMBL: AF122922; AAD25402.1; -.  
 DR HSSP: P00743; IICF.  
 DR INTERPRO: IPR000561; -.  
 DR INTERPRO: IPR002049; -.  
 DR PFAM: PF00008; EGF. 5.  
 DR PRINTS: PR00011; EGF\_LAMININ.  
 DR PROSITE: PS00022; EGF\_1; UNKNOWN\_5.  
 DR PROSITE: PS01186; EGF\_2; 4.  
 KW Repeat; EGF-like domain; Signal; Developmental protein.  
 FT CHAIN 1 28  
 FT SIGNAL 1 28  
 FT DOMAIN 29 379  
 FT DOMAIN 177 208  
 FT DOMAIN 209 240  
 FT DOMAIN 241 272  
 FT DOMAIN 273 304  
 FT DOMAIN 305 336  
 FT DOMAIN 177 186  
 FT DISULFID 182 192  
 FT DISULFID 198 208  
 FT DISULFID 209 218  
 FT DISULFID 214 224  
 FT DISULFID 220 232  
 FT DISULFID 241 250  
 FT DISULFID 246 256  
 FT DISULFID 262 264  
 FT DISULFID 273 282  
 FT DISULFID 278 288  
 FT DISULFID 294 296  
 FT DISULFID 305 314  
 FT DISULFID 310 320  
 FT DISULFID 326 328  
 SQ SEQUENCE 379 AA; 41512 MW; 27782370A266E784 CRC64;

Query Match 87.2%; Score 41; DB 4; Length 379;  
 Best Local Similarity 87.5%; Pred. No. 4.3;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GGLCVOPG 8  
 DB 221 GGLCVTPG 228

RESULT 4  
 O9WUAI PRELIMINARY; PRT; 379 AA.  
 ID Q9WUAI;  
 AC Q9WUAI;  
 DT 01-NOV-1999 (TRENBLREL. 12, Created)  
 DT 01-NOV-1999 (TRENBLREL. 12, Last sequence update)  
 DT 01-OCT-2000 (TRENBLREL. 15, Last annotation update)  
 DE WNT INHIBITORY FACTOR 1 PRECURSOR (WNT INHIBITORY FACTOR-1) (WIF-1).  
 GN WIF1 OR WIF-1.



```

OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99215557; PubMed=10201374;
RA Hsieh J.-C., Kodjabachian L., Rebert M.L., Ratner A.,
RA Smallwood P.M., Samos C.H., Nusse R., David I.B., Nathans J.;
RT "A new secreted protein that binds to Wnt proteins and inhibits their
RT activities."
RL Nature 398:431-436(1999).
CC -1- FUNCTION: BINDS TO WNT PROTEINS AND INHIBITS THEIR ACTIVITIES. MAY
CC BE INVOLVED IN MESODERM SEGMENTATION.
CC -1- TISSUE SPECIFICITY: EXPRESSION HIGHEST IN HEART AND LUNG. LOWER IN
CC BRAIN AND EYE.
CC -1- SIMILARITY: CONTAINS 5 EGF-LIKE DOMAINS.
DR EMBL: AF122923; AAD25403.1; -.
DR HSSP: P00740; 11XA.
DR MGD: MGI:1344332; Wt1f1-pending.
DR INTERPRO: IPR000561; -.
DR PFM: PF00008; EGF_5.
DR PRINTS: PR00011; EGF_LAMININ.
DR PROSITE: PS00022; EGF_1; UNKNOWN_5.
DR PROSITE: PS01186; EGF_2; 4.
KW Repeat; EGF-like domain; Signal; Developmental protein.
FT SIGNAL 1 28
FT CHAIN 29 379 WNT INHIBITORY FACTOR 1.
FT DOMAIN 177 208 EGF-LIKE 1.
FT DOMAIN 209 240 EGF-LIKE 2.
FT DOMAIN 241 272 EGF-LIKE 3.
FT DOMAIN 273 304 EGF-LIKE 4.
FT DOMAIN 305 336 EGF-LIKE 5.
FT DISULFID 177 186 POTENTIAL.
FT DISULFID 182 192 POTENTIAL.
FT DISULFID 198 200 POTENTIAL.
FT DISULFID 209 218 POTENTIAL.
FT DISULFID 214 224 POTENTIAL.
FT DISULFID 230 232 POTENTIAL.
FT DISULFID 241 250 POTENTIAL.
FT DISULFID 246 256 POTENTIAL.
FT DISULFID 262 264 POTENTIAL.
FT DISULFID 273 282 POTENTIAL.
FT DISULFID 278 288 POTENTIAL.
FT DISULFID 294 296 POTENTIAL.
FT DISULFID 305 314 POTENTIAL.
FT DISULFID 310 320 POTENTIAL.
FT DISULFID 326 328 POTENTIAL.
SQ SEQUENCE 379 AA; 41590 MW; E3765F2642B2BC9A CRC64;

Query Match 87.2%; Score 41; DB 11; Length 379;
Best Local Similarity 87.5%; Pred. No. 4.3;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 1 GGLCYVPG 8
DB 221 GGLCVTPG 228

```

```

OC Cypriniformes; Cyprinidae; Rasbora; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99215557; PubMed=10201374;
RA Hsieh J.-C., Kodjabachian L., Rebert M.L., Ratner A.,
RA Smallwood P.M., Samos C.H., Nusse R., David I.B., Nathans J.;
RT "A new secreted protein that binds to Wnt proteins and inhibits their
RT activities."
RL Nature 398:431-436(1999).
CC -1- FUNCTION: BINDS TO WNT PROTEINS AND INHIBITS THEIR ACTIVITIES. MAY
CC BE INVOLVED IN MESODERM SEGMENTATION.
CC -1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN UNSEGMENTED PARAXIAL
CC MESODERM.
CC -1- SIMILARITY: CONTAINS 5 EGF-LIKE DOMAINS.
DR EMBL: AF122925; AAD25405.1; -.
DR HSSP: P00740; 1EDM.
DR ZFIN: ZDB-GENE-990712-17; wtf1.
DR INTERPRO: IPR000561; -.
DR PFM: PF00008; EGF_5.
DR PRINTS: PR00011; EGF_LAMININ.
DR PROSITE: PS00022; EGF_1; UNKNOWN_5.
DR PROSITE: PS01186; EGF_2; 4.
KW Repeat; EGF-like domain; Signal; Developmental protein.
FT SIGNAL 1 28
FT CHAIN 29 378 WNT INHIBITORY FACTOR 1.
FT DOMAIN 175 206 EGF-LIKE 1.
FT DOMAIN 207 238 EGF-LIKE 2.
FT DOMAIN 239 270 EGF-LIKE 3.
FT DOMAIN 271 302 EGF-LIKE 4.
FT DOMAIN 303 334 EGF-LIKE 5.
FT DISULFID 175 184 POTENTIAL.
FT DISULFID 180 190 POTENTIAL.
FT DISULFID 196 198 POTENTIAL.
FT DISULFID 207 216 POTENTIAL.
FT DISULFID 212 222 POTENTIAL.
FT DISULFID 228 230 POTENTIAL.
FT DISULFID 239 248 POTENTIAL.
FT DISULFID 244 254 POTENTIAL.
FT DISULFID 260 262 POTENTIAL.
FT DISULFID 271 280 POTENTIAL.
FT DISULFID 276 286 POTENTIAL.
FT DISULFID 292 294 POTENTIAL.
FT DISULFID 303 312 POTENTIAL.
FT DISULFID 308 318 POTENTIAL.
FT DISULFID 324 326 POTENTIAL.
SQ SEQUENCE 378 AA; 41312 MW; 42FE9F70D948D1D8 CRC64;

Query Match 83.0%; Score 39; DB 13; Length 378;
Best Local Similarity 75.0%; Pred. No. 10;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
OY 1 GGLCYVPG 8
DB 219 GGLCMSPG 226

```

```

RESULT 5
ID 09W6F9 PRELIMINARY; PRT; 378 AA.
AC 09W6F9;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE WNT INHIBITORY FACTOR 1 PRECURSOR (WNT INHIBITORY FACTOR-1) (WIF-1).
GN WIF1 OR WIF-1.
OS Brachydanio rerio (Zebrafish) (Zebra danio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;

```

```

RESULT 6
ID 075851 PRELIMINARY; PRT; 4123 AA.
AC 075851;
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE WUSC-H.D10751H13.1 PROTEIN (FRAGMENT)..
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.

```

```

RA Leonard S., Graves T., Strommatt C.;
RT "The sequence of Homo sapiens PAC clone D10751H13."
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Waterston R.H.;
RN Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Waterston R.;
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC004877; AAC36301.1; -
DR HSSP; P01130; IAJJ.
DR INTERPRO; IPRO00421; -
DR INTERPRO; IPRO00561; -
DR INTERPRO; IPRO00884; -
DR INTERPRO; IPRO00923; -
DR INTERPRO; IPRO01007; -
DR INTERPRO; IPRO01064; -
DR INTERPRO; IPRO01092; -
DR INTERPRO; IPRO01846; -
DR INTERPRO; IPRO02172; -
DR INTERPRO; IPRO02223; -
DR INTERPRO; IPRO02465; -
DR INTERPRO; IPRO02919; -
DR PFAM; PF00057; 1d1_recept_a; 11.
DR PFAM; PF00090; tsp_1; 14.
DR PFAM; PF00094; vwd; 5.
DR PFAM; PF00754; F5_F8_type_C; 1.
DR PRNTS; P01826; TIL; 9.
DR PROSITE; PRO0261; LDLRECEPTOR.
DR PROSITE; PS00022; EGF_L1; UNKNOWN_1.
DR PROSITE; PS00196; COPPER_BLDG; UNKNOWN_1.
DR PROSITE; PS00225; CRYSTALLIN_BETAGAMMA; UNKNOWN_1.
DR PROSITE; PS00280; BPTI_KUNITZ; UNKNOWN_1.
DR PROSITE; PS00340; RECEPTOR_CYTOKINES_2; UNKNOWN_6.
DR PROSITE; PS01209; LDLRA_1; 9.
DR PROSITE; PS50068; LDLRA_2; 20.
KM Glycoprotein.
FT NON_TER
SQ SEQUENCE 4123 AA; 434985 MW; 7AAB6F8DCE012FB CRC64;

Query Match
Best Local Similarity 83.0%; Score 39; DB 4; Length 4123;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGLCVOPG 8
Db 3234 GGLCVOPG 3241

```

```

FT NON_TER
SQ SEQUENCE 368 AA; 40584 MW; 1C540BD58DD9D3715 CRC64;

Query Match
Best Local Similarity 80.9%; Score 38; DB 11; Length 368;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGLCVOP 7
Db 81 GGLCVOP 87

RESULT 8
ID 075543 PRELIMINARY; PRT; 361 AA.
AC 075543;
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT 01-AUG-1999 (TREMBLrel. 11, Last annotation update)
DE HYPOTHETICAL 41.9 KDA PROTEIN (FRAGMENT).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN;
RX MEDLINE-96207227; PubMed-8619474;
RA Andersson B., Wentland M.A., Ricafrente J.Y., Liu W., Gibbs R.A.;
RT "A 'double adaptor' method for improved shotgun library
construction."
RL Anal. Biochem. 236:107-113(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN;
RX MEDLINE-97264341; PubMed-9110174;
RA Yu W., Andersson B., Worley K.C., Muzny D.M., Ding Y., Liu W.,
RA Ricafrente J.Y., Wentland M.A., Lennon G., Gibbs R.A.;
RT "Large-scale concatenation cDNA sequencing."
RL Genome Res. 7:353-358(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN;
RA Yu W., Gibbs R.A.;
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF070614; AAC25386.1; -
KM Hypothetical protein.
FT NON_TER
SQ SEQUENCE 381 AA; 41902 MW; 72F589485A571CBF CRC64;

Query Match
Best Local Similarity 80.9%; Score 38; DB 4; Length 381;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGLCVOP 7
Db 94 GGLCVOP 100

RESULT 9
ID 09P0W3 PRELIMINARY; PRT; 474 AA.
AC 09P0W3;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE SCHIP-1-D241/253.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;

```

[1]  
 RN SEQUENCE FROM N.A.  
 RP TISSUE-BRAIN:  
 RA Goudebroze L., Brault E., Muchardt C., Camonis J., Thomas G.:  
 RT "Cloning and Characterization of SCHIP-1, a Novel Protein Interacting  
 RT Specifically with Spliced Isoforms and Naturally Occurring Mutant NF2  
 RT Proteins".  
 RL Mol. Cell. Biol. 20:1699-1712(2000).  
 DR EMBL: AF145715; AAF34243.1; -  
 SQ SEQUENCE 474 AA; 52079 MW; 5CAE98E1867A78F9 CRC64;

Query Match 80.9%; Score 38; DB 4; Length 474;  
 Best Local Similarity 85.7%; Pred. No. 20;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GGLCVOP 7  
 |||||:  
 DB 200 GGLCLDP 206

RESULT 10  
 O9P0W5 PRELIMINARY; PRT; 487 AA.  
 AC O9P0W5;  
 DT 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)  
 DE SCHIP-1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-BRAIN:  
 RA Goudebroze L., Brault E., Muchardt C., Camonis J., Thomas G.:  
 RT "Cloning and Characterization of SCHIP-1, a Novel Protein Interacting  
 RT Specifically with Spliced Isoforms and Naturally Occurring Mutant NF2  
 RT Proteins".  
 RL Mol. Cell. Biol. 20:1699-1712(2000).  
 DR EMBL: AF145713; AAF34241.1; -  
 SQ SEQUENCE 487 AA; 53479 MW; 86F86EDD64A75BF CRC64;

Query Match 80.9%; Score 38; DB 4; Length 487;  
 Best Local Similarity 85.7%; Pred. No. 20;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GGLCVOP 7  
 |||||:  
 DB 200 GGLCLDP 206

RESULT 11  
 O02660 PRELIMINARY; PRT; 870 AA.  
 AC O02660;  
 DT 01-JUL-1997 (TREMBLrel. 04, Created)  
 DT 01-JAN-1999 (TREMBLrel. 09, Last sequence update)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update).  
 DE SCO-SPONDIN (FRAGMENT).  
 OS Bos taurus (Bovine).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 OC Bovidae; Bovinae; Bos.  
 OX NCBI\_TaxID=9913;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-SUBCOMMISSURAL ORGAN;  
 RA Godron S.:  
 RT Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: Y08560; CAA69867.1; -

DR HSP: P01130; IAUJ.  
 DR INTERPRO: IPR000884; -  
 DR INTERPRO: IPR001846; -  
 DR INTERPRO: IPR002172; -  
 DR INTERPRO: IPR002465; -  
 DR INTERPRO: IPR002919; -  
 DR PFAM: PF000057; 1d1\_recept\_a; 7.  
 DR PFAM: PF00090; tsp\_1; 1.  
 DR PFAM: PF00094; vwd; 1.  
 DR PFAM: PF01826; tll; 2.  
 DR PRINTS: PR00261; LDLRECEPTOR.  
 DR PROSITE: PS00340; RECEPTOR\_CYTOKINES\_2; UNKNOWN\_2.  
 DR PROSITE: PS01209; LDLRA\_1; 4.  
 DR PROSITE: PS50068; LDLRA\_2; 6.  
 KM Glycoprotein.  
 FT NON\_TER 1  
 FT NON\_TER 870  
 SQ SEQUENCE 870 AA; 91875 MW; DD9D4F7B77AC48CE CRC64;

Query Match 80.9%; Score 38; DB 6; Length 870;  
 Best Local Similarity 75.0%; Pred. No. 35;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 GGLCVOP 8  
 |||||:  
 DB 474 GGRCVEPG 481

RESULT 12  
 O03515 PRELIMINARY; PRT; 250 AA.  
 AC O03515;  
 DT 01-JAN-1998 (TREMBLrel. 05, Created)  
 DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)  
 DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)  
 DE HYPOTHETICAL PROTEIN.  
 OS Rhodospirillum rubrum (Rhodospirillum rubrum).  
 OC Bacteria; Proteobacteria; alpha subdivision; Rhodospirillum group;  
 OC Rhodospirillum.  
 OX NCBI\_TaxID=1061;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C.B. VAN NIEL ATH 2.3.1, ATCC CATALOG #11166;  
 RA Larimer F.W., Lu T.-Y.S., Buile D.M.;  
 RL FASEB J. 9:0-0(1997).  
 DR EMBL: U23145; AAB82054.1; -  
 DR INTERPRO: IPR000792; -  
 DR PFAM: PF00196; Gerd; 1.  
 SQ SEQUENCE 250 AA; 27505 MW; 69C87294604476F9 CRC64;

Query Match 78.7%; Score 37; DB 2; Length 250;  
 Best Local Similarity 85.7%; Pred. No. 17;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 GGLCVOP 8  
 |||||:  
 DB 138 GGLACPG 144

RESULT 13  
 O90XT5 PRELIMINARY; PRT; 278 AA.  
 AC O90XT5;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)  
 DE NOTCH4-LIKE PROTEIN.  
 GN ZNEU1.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Sheppard P., Jelinek L., Whitmore T., Blumberg H., Lehner J.,  
 RA O'Hara P.;  
 RT "Mus musculus homolog of HLA class III region containing NOTCH4  
 gene";  
 RL Submitted (OCT-1999) to the EMBL/Genbank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Pressnell S., Gilbert T., Whitmore T., Foster D., Hart C., Lehner J.,  
 RA Matinez T., Hoffman R., O'Hara P.;  
 RL Submitted (SEP-1999) to the EMBL/Genbank/DBJ databases.  
 DR EMBL; AF184973; AAF01332.1; -  
 DR HSSP; P00740; IEDM.  
 DR INTERPRO; IPR000561; -  
 DR INTERPRO; IPR001881; -  
 DR PRAM; PF00008; EGF\_2.  
 DR PROSITE; PS00022; EGF\_1; UNKNOWN\_1.  
 DR PROSITE; PS01186; EGF\_2; UNKNOWN\_1.  
 DR PROSITE; PS01187; EGF\_CA; 1.  
 SQ SEQUENCE 278 AA; 30125 MW; 863159f567102f7a7 CRC64;

Query Match 76.6%; Score 36; DB 11; Length 278;  
 Best Local Similarity 62.5%; Pred. No. 28;  
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 GGLCVOPG 8  
 |||111111  
 Db 118 GGACSQPG 125

RESULT 14  
 OY 094106 PRELIMINARY; PRT; 645 AA.  
 AC 094106;  
 DT 01-MAY-1999 (TREMBlrel. 10, Created)  
 DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)  
 DE 01-OCT-2000 (TREMBlrel. 15, Last annotation update)  
 DE HEAT SHOCK PROTEIN 70 (FRAGMENT).  
 GN PCSA1.  
 OS Pneumocystis carinii f. sp. carinii.  
 OC Eukaryota; Fungi; Fungi Incertae sedis; Pneumocystidaceae;  
 OC Pneumocystis.  
 OX NCBI\_TaxID=36081;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-97045128; Pubmed-8890193;  
 RA Steadman T.T., Buck G.A.;  
 RT "Identification, characterization, and expression of the Bip  
 endoplasmic reticulum resident chaperonins in Pneumocystis carinii.";  
 RL Infect. Immun. 64:4463-4471(1996).  
 DR EMBL; U80967; AAD00455.1; -  
 DR HSSP; P19120; 3HSC.  
 DR INTERPRO; IPR000169; -  
 DR INTERPRO; IPR001023; -  
 DR PRAM; PF000012; HSP70; 1.  
 DR PRINTS; PR00301; HEATSHOCK70.  
 DR PROSITE; PS00287; HSP70\_1; UNKNOWN\_1.  
 DR PROSITE; PS00329; HSP70\_2; 1.  
 DR PROSITE; PS00639; THIOI\_PROTEASE\_HIS; UNKNOWN\_1.  
 DR PROSITE; PS01036; HSP70\_3; 1.  
 KW Heat shock.  
 FT NON\_TER  
 SQ SEQUENCE 645 AA; 70884 MW; 6D8CF90433BBI63F CRC64;

Query Match 76.6%; Score 36; DB 3; Length 645;  
 Best Local Similarity 75.0%; Pred. No. 62;  
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 OY 1 GGLCVOPG 8

Db 621 GGACSQPG 628

RESULT 15  
 OY 094104 PRELIMINARY; PRT; 647 AA.  
 AC 094104;  
 DT 01-MAY-1999 (TREMBlrel. 10, Created)  
 DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)  
 DE 01-OCT-2000 (TREMBlrel. 15, Last annotation update)  
 DE HEAT SHOCK PROTEIN 70 (FRAGMENT).  
 GN HSP70.  
 OS Pneumocystis carinii.  
 OC Eukaryota; Fungi; Fungi Incertae sedis; Pneumocystidaceae;  
 OC Pneumocystis.  
 OX NCBI\_TaxID=4754;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Paul S.P., Graves D.C.;  
 RT "Phylogeny and sequence analysis of Pneumocystis carinii HSP70.";  
 RL Submitted (SEP-1996) to the EMBL/Genbank/DBJ databases.  
 DR EMBL; U71151; AAD09565.1; -  
 DR HSSP; P19120; 3HSC.  
 DR INTERPRO; IPR000169; -  
 DR INTERPRO; IPR001023; -  
 DR PRAM; PF000012; HSP70; 1.  
 DR PRINTS; PR00301; HEATSHOCK70.  
 DR PROSITE; PS00287; HSP70\_1; UNKNOWN\_1.  
 DR PROSITE; PS00329; HSP70\_2; 1.  
 DR PROSITE; PS00639; THIOI\_PROTEASE\_HIS; UNKNOWN\_1.  
 DR PROSITE; PS01036; HSP70\_3; 1.  
 KW Heat shock.  
 FT NON\_TER  
 SQ SEQUENCE 647 AA; 71176 MW; 92F94963999380F9 CRC64;

Query Match 76.6%; Score 36; DB 3; Length 647;  
 Best Local Similarity 75.0%; Pred. No. 63;  
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 GGLCVOPG 8  
 |||111111  
 Db 623 GGACSQPG 630

Search completed: June 13, 2001, 14:20:36  
 Job time: 736 sec



*This Page Blank (uspto)*

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

## OM protein - protein search, using sw model

Run on: June 13, 2001, 14:16:36 ; Search time 118.55 Seconds  
(without alignments)  
1.296 Million cell updates/sec

Title: PCT-US01-05825A-17  
Perfect score: 47  
Sequence: 1 GGLCVQPG 8

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 185757 seqs, 19210857 residues

Total number of hits satisfying chosen parameters: 185757

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 08  
Maximum Match 1008  
Listing first 45 summaries

Database : Issued Patents, AA: \*  
1: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep:\*  
2: /cgn2\_6/ptodata/2/1aa/5B.COMB.pep:\*  
3: /cgn2\_6/ptodata/2/1aa/6A.COMB.pep:\*  
4: /cgn2\_6/ptodata/2/1aa/6B.COMB.pep:\*  
5: /cgn2\_6/ptodata/2/1aa/PCITUS.COMB.pep:\*  
6: /cgn2\_6/ptodata/2/1aa/Backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	35	74.5	531	2	US-08-975-114A-4
2	35	74.5	531	4	US-08-849-281A-4
3	35	74.5	536	2	US-08-975-114A-5
4	35	74.5	536	4	US-08-849-281A-3
5	34	72.3	20	2	US-08-859-931A-4
6	34	72.3	118	3	US-08-545-809A-125
7	34	72.3	130	4	US-09-347-819-8
8	34	72.3	2556	1	US-08-185-432-17
9	34	72.3	2556	1	US-08-083-550A-20
10	34	72.3	2556	3	US-08-533-384-20
11	33	70.2	489	1	US-08-434-702-4
12	32	68.1	118	3	US-08-545-809A-108
13	32	68.1	123	5	US-08-428-197-38
14	32	68.1	123	5	PCT-US93-10555-38
15	32	68.1	866	3	US-08-651-999A-1
16	32	68.1	890	1	US-08-445-640-2
17	32	68.1	890	3	US-08-170-558-2
18	32	68.1	890	3	US-08-447-314-2
19	32	68.1	890	3	US-08-445-461-2
20	32	68.1	911	1	US-08-286-305A-1
21	32	68.1	911	2	US-08-441-104A-1
22	32	68.1	911	2	US-08-440-816A-1
23	32	68.1	968	3	US-08-651-999A-7
24	32	68.1	1239	4	US-08-937-931-2
25	32	68.1	1239	4	US-09-285-502-2
26	31	66.0	92	2	US-07-728-215-39
27	31	66.0	92	2	US-07-728-215-41

28	31	66.0	120	1	US-08-478-039-101	Sequence 101, App
29	31	66.0	120	1	US-08-476-349A-101	Sequence 101, App
30	31	66.0	383	1	US-08-597-545-2	Sequence 2, Appl1
31	31	66.0	383	1	US-08-457-135-2	Sequence 2, Appl1
32	31	66.0	400	2	US-08-624-601-8	Sequence 8, Appl1
33	31	66.0	573	2	US-08-745-934-4	Sequence 4, Appl1
34	31	66.0	573	4	US-09-147-009-7	Sequence 7, Appl1
35	31	66.0	718	1	US-08-444-792-4	Sequence 4, Appl1
36	31	66.0	718	1	US-08-445-042-4	Sequence 4, Appl1
37	31	66.0	753	4	US-08-942-686-2	Sequence 2, Appl1
38	31	66.0	788	2	US-07-728-215-32	Sequence 32, Appl1
39	30	63.8	119	2	US-08-475-000-16	Sequence 16, Appl1
40	30	63.8	119	2	US-08-483-199-16	Sequence 16, Appl1
41	30	63.8	119	2	US-08-484-508-16	Sequence 16, Appl1
42	30	63.8	142	1	US-08-681-812-2	Sequence 2, Appl1
43	30	63.8	150	4	US-09-188-930-306	Sequence 306, App
44	30	63.8	207	2	US-08-808-550-36	Sequence 36, Appl1
45	30	63.8	212	1	US-08-461-859-35	Sequence 35, Appl1

## ALIGNMENTS

RESULT 1  
US-08-975-114A-4  
Sequence 4, Application US/08975114A  
Patent No. 5876714  
GENERAL INFORMATION:  
APPLICANT: Acsush1 NISHIKAWA et al.  
TITLE OF INVENTION: HUMAN GLYCOSYLTRANSFERASE GENE.  
TITLE OF INVENTION: COMPOUNDS AND METHOD FOR INHIBITTING CANCEROUS METASTASIS  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Wenderoth, Lind & Ponack, L.L.P.  
STREET: 2033 K Street, N.W., Suite 800  
CITY: Washington  
STATE: D.C.  
COUNTRY: U.S.A.  
ZIP: 20006  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: Wordperfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/975,114A  
FILING DATE: No. 5876714ember 20, 1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/107,173  
FILING DATE: August 17, 1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Lee Cheng  
REGISTRATION NUMBER: 40,949  
REFERENCE/DOCKET NUMBER: 1-E3439DIY  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-721-8200  
TELEFAX: 202-721-8250  
TELEX:  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 531 amino acid residues  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE:  
US-08-975-114A-4

Query Match 74.5%; Score 35; DB 2; Length 531;  
Best Local Similarity 62.5%; Pred. No. 1.4e+02;  
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 GGLCVOPG 8  
11:1:11  
Db 112 GGVCFKPG 119

RESULT 2  
US-08-849-281A-4  
; Sequence 4, Application US/08849281A  
; Patent No. 6153433  
; GENERAL INFORMATION:  
; APPLICANT: Eiji MIYOSHI et al.  
; TITLE OF INVENTION: INHIBITOR FOR VIRAL REPLICATION  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Wenderoth, Lind & Ponack, L.L.P.  
; STREET: 2033 K Street, N.W., Suite 800  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: U.S.A.  
; ZIP: 20006  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: MS-DOS  
; SOFTWARE: Wordperfect 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/849,281A  
; FILING DATE: May 30, 1997  
; CLASSIFICATION: 514  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Lee Cheng  
; REGISTRATION NUMBER: 40,949  
; REFERENCE/DOCKET NUMBER: 97-0529\*/LC(MJ)/1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-721-8200  
; TELEFAX: 202-721-8250  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 531 amino acid residues  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: polypeptide  
US-08-849-281A-4

Query Match 74.5%; Score 35; DB 4; Length 531;  
Best Local Similarity 62.5%; Pred. No. 1.4e+02;  
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 GGLCVOPG 8  
11:1:11  
Db 112 GGVCFKPG 119

RESULT 3  
US-08-975-114A-5  
; Sequence 5, Application US/08975114A  
; Patent No. 5876714  
; GENERAL INFORMATION:  
; APPLICANT: Atsushi NISHIKAWA et al.  
; TITLE OF INVENTION: HUMAN GLYCOSYLTRANSFERASE GENE,  
; TITLE OF INVENTION: COMPOUNDS AND METHOD FOR INHIBITING CANCEROUS METASTASIS  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Wenderoth, Lind & Ponack, L.L.P.  
; STREET: 2033 K Street, N.W., Suite 800  
; CITY: Washington  
; STATE: D.C.

COUNTRY: U.S.A.  
ZIP: 20006  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: Wordperfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/975,114A  
FILING DATE: No. 5876714ember 20, 1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/107,173  
FILING DATE: August 17, 1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Lee Cheng  
REGISTRATION NUMBER: 40,949  
REFERENCE/DOCKET NUMBER: 1-F3439DIV  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-721-8200  
TELEFAX: 202-721-8250  
TELEX:  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 536 amino acid residues  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE:  
US-08-975-114A-5

Query Match 74.5%; Score 35; DB 2; Length 536;  
Best Local Similarity 62.5%; Pred. No. 1.4e+02;  
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 GGLCVOPG 8  
11:1:11  
Db 112 GGVCFKPG 119

RESULT 4  
US-08-849-281A-3  
; Sequence 3, Application US/08849281A  
; Patent No. 6153433  
; GENERAL INFORMATION:  
; APPLICANT: Eiji MIYOSHI et al.  
; TITLE OF INVENTION: INHIBITOR FOR VIRAL REPLICATION  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Wenderoth, Lind & Ponack, L.L.P.  
; STREET: 2033 K Street, N.W., Suite 800  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: U.S.A.  
; ZIP: 20006  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: MS-DOS  
; SOFTWARE: Wordperfect 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/849,281A  
; FILING DATE: May 30, 1997  
; CLASSIFICATION: 514  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Lee Cheng  
; REGISTRATION NUMBER: 40,949  
; REFERENCE/DOCKET NUMBER: 97-0529\*/LC(MJ)/1  
; TELECOMMUNICATION INFORMATION:



TELEPHONE: 202-721-8200  
TELEFAX: 202-721-8250  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 536 amino acid residues  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: polypeptide  
US-08-849-281A-3

Query Match 74.5%; Score 35; DB 4; Length 536;  
Best Local Similarity 62.5%; Pred. No. 1.4e+02;  
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGLCVOPG 8  
DB 112 GGVCFRPG 119

RESULT 5  
US-08-859-931A-4  
Sequence 4, Application US/0885931A  
Patent No. 5945510  
GENERAL INFORMATION:  
APPLICANT: FASANO, Alessio  
TITLE OF INVENTION: SUBSTANTIALLY PURE ZONULIN, A  
TITLE OF INVENTION: PHYSIOLOGICAL MODULATOR OF  
TITLE OF INVENTION: MAMMALIAN TIGHT JUNCTIONS  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SUGHRUE, MION, ZINN, MACPEAK & SEAS  
STREET: 2100 Pennsylvania Avenue, N.W., Suite 800  
CITY: Washington, D.C.  
STATE: D.C.  
COUNTRY: U.S.A.  
ZIP: 20037  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/859,931A  
FILING DATE: 21 MAY 1997  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: KIT, Gordon  
REGISTRATION NUMBER: 30,764  
REFERENCE/DOCKET NUMBER: A-6901  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 293-7060  
TELEFAX: (202) 293-7860  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 20 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
HYPOTHETICAL: NO  
US-08-859-931A-4

Query Match 72.3%; Score 34; DB 2; Length 20;  
Best Local Similarity 75.0%; Pred. No. 8.7;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGLCVOPG 8  
DB 8 GGVLVOPG 15

RESULT 6  
US-08-545-809A-125  
Sequence 125, Application US/08545809A  
Patent No. 6096878  
GENERAL INFORMATION:  
APPLICANT: Hoojo, Tasuku  
APPLICANT: Matsuda, Fumihiko  
TITLE OF INVENTION: HUMAN IMMUNOGLOBULIN VH GENE  
TITLE OF INVENTION: SEGMENTS AND DNA FRAGMENTS CONTAINING THE SAME  
NUMBER OF SEQUENCES: 145  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson, P.C.  
STREET: 225 Franklin Street  
CITY: Boston  
STATE: MA  
COUNTRY: US  
ZIP: 02110-2804  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: Windows95  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/545,809A  
FILING DATE: 27-MAR-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/JP93/00603  
FILING DATE: 10-MAY-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Freeman, John W.  
REGISTRATION NUMBER: 29,066  
REFERENCE/DOCKET NUMBER: 06501/004001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-542-5070  
TELEFAX: 617-542-8906  
TELEX: 200154  
INFORMATION FOR SEQ ID NO: 125:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 118 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-545-809A-125

Query Match 72.3%; Score 34; DB 3; Length 118;  
Best Local Similarity 75.0%; Pred. No. 48;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGLCVOPG 8  
DB 27 GGVLVOPG 34

RESULT 7  
US-09-347-819-8  
Sequence 8, Application US/09347819  
Patent No. 6184036  
GENERAL INFORMATION:  
APPLICANT: Cahoon, Rebecca E.  
APPLICANT: Rafalski, J. Antoni  
APPLICANT: Falco, S. Carl  
TITLE OF INVENTION: Ornithine Biosynthesis Enzymes  
FILE REFERENCE: BB-1174-C  
CURRENT APPLICATION NUMBER: US/09/347,819  
CURRENT FILING DATE: 1999-07-02  
EARLIER APPLICATION NUMBER: 60/093,209  
EARLIER FILING DATE: July 17, 1998  
NUMBER OF SEQ ID NOS: 14  
SOFTWARE: Microsoft Office 97  
SEQ ID NO 8  
LENGTH: 130

TYPE: PRT  
ORGANISM: Glycine max  
FEATURE:  
NAME/KEY: UNSURE  
LOCATION: (109)  
US-09-347-819-8

Query Match 72.3%; Score 34; DB 4; Length 130;  
Best Local Similarity 83.3%; Pred. No. 53;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 LCVQPG 8  
111111  
DB 74 LCIOPG 79

RESULT 8  
US-08-185-432-17  
Sequence 17, Application US/08185432  
Patent No. 5750652  
GENERAL INFORMATION:  
APPLICANT: Artavanis-Tsakonas, Spyridon  
APPLICANT: Busseau, Isabelle  
APPLICANT: Diederich, Robert J.  
APPLICANT: Xu, Tian  
APPLICANT: Matsuno, Kenji  
TITLE OF INVENTION: DELTEX PROTEINS, NUCLEIC ACIDS, AND  
TITLE OF INVENTION: ANTIBODIES, AND RELATED METHODS AND COMPOSITIONS  
NUMBER OF SEQUENCES: 23  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: PENNIE & EDMONDS  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036-2711  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/185,432  
FILING DATE: 21-JAN-1994  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: Mistock, S. Leslie  
REGISTRATION NUMBER: 18,872  
REFERENCE/DOCKET NUMBER: 7326-006  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-8864/9741  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 17:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2556 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
US-08-185-432-17

Query Match 72.3%; Score 34; DB 1; Length 2556;  
Best Local Similarity 75.0%; Pred. No. 9.1e+02;  
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GGLCVQPG 8  
111111  
DB 1120 GGLCVQDAG 1127

RESULT 9

US-08-083-590A-20  
Sequence 20, Application US/08083590A  
Patent No. 5786158  
GENERAL INFORMATION:  
APPLICANT: Artavanis-Tsakonas, S. et al.  
TITLE OF INVENTION: Therapeutic And Diagnostic Methods  
TITLE OF INVENTION: And Compositions Based On No. 5786158ch Proteins And  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pennie & Edmonds  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/083,590A  
FILING DATE: 25-JUN-1993  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Mistock, S. Leslie  
REGISTRATION NUMBER: 18,872  
REFERENCE/DOCKET NUMBER: 7326-015  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212 790-9090  
TELEFAX: 212 8698864/9741  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 20:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2556 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
US-08-083-590A-20

Query Match 72.3%; Score 34; DB 1; Length 2556;  
Best Local Similarity 75.0%; Pred. No. 9.1e+02;  
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GGLCVQPG 8  
111111  
DB 1120 GGLCVQDAG 1127

RESULT 10  
US-08-532-384-20  
Sequence 20, Application US/08532384  
Patent No. 6083904  
GENERAL INFORMATION:  
APPLICANT: Artavanis-Tsakonas, S. et al.  
TITLE OF INVENTION: Therapeutic And Diagnostic Methods  
TITLE OF INVENTION: And Compositions Based On No. 6083904ch Proteins And  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pennie & Edmonds  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/532,384  
FILING DATE:  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/083,590  
FILING DATE: 25-JUN-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: MIBROCK, S. Leslie  
REGISTRATION NUMBER: 18,872  
REFERENCE/DOCKET NUMBER: 7326-015  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212 790-9090  
TELEFAX: 212 8698864/9741  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 20:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2556 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
US-08-532-384-20

Query Match 72.3%; Score 34; DB 3; Length 2556;  
Best Local Similarity 75.0%; Pred. No. 9.1e+02;  
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GGLCVOPG 8  
||| |  
DB 1120 GGLCVDPAG 1127

RESULT 11  
US-08-434-702-4  
Sequence 4, Application US/08434702  
Patent No. 5554743  
GENERAL INFORMATION:  
APPLICANT: Bennett, Alan B.  
APPLICANT: Fischer, Robert L.  
APPLICANT: Lashbrook, Coralie  
APPLICANT: Giovannoli, James  
TITLE OF INVENTION: Endo-1,4-beta-glucanase Genes and Their  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend and Crew  
STREET: One Market Plaza, Stuart Street Tower  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94105-1492  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/434,702  
FILING DATE: 04-MAY-1995  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/271,883  
FILING DATE: 07-JUL-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/687,466  
FILING DATE: 18-APR-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/511,417  
FILING DATE: 20-APR-1990  
ATTORNEY/AGENT INFORMATION:

NAME: Bastian, Kevin L.  
REGISTRATION NUMBER: 34,774  
REFERENCE/DOCKET NUMBER: 02307E-304300S  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 543-9600  
TELEFAX: (415) 543-5043  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 489 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-434-702-4

Query Match 70.2%; Score 33; DB 1; Length 489;  
Best Local Similarity 75.0%; Pred. No. 2.7e+02;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGLCVOPG 8  
||| |  
DB 327 GGLIVKPG 334

RESULT 12  
US-08-545-809A-108  
Sequence 108, Application US/08545809A  
Patent No. 6096878  
GENERAL INFORMATION:  
APPLICANT: Honjo, Tasuku  
APPLICANT: Matsuda, Fumihiko  
TITLE OF INVENTION: HUMAN IMMUNOGLOBULIN VH GENE  
SEGMENTS AND DNA FRAGMENTS CONTAINING THE SAME  
NUMBER OF SEQUENCES: 145  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson, P.C.  
STREET: 225 Franklin Street  
CITY: Boston  
STATE: MA  
COUNTRY: US  
ZIP: 02110-2804  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: Windows95  
SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/545,809A  
FILING DATE: 27-MAR-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/JP93/00603  
FILING DATE: 10-MAY-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Freeman, John W.  
REGISTRATION NUMBER: 29,066  
REFERENCE/DOCKET NUMBER: 06501/004001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-542-5070  
TELEFAX: 617-542-8906  
TELEX: 200154  
INFORMATION FOR SEQ ID NO: 108:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 118 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-545-809A-108

Query Match 68.1%; Score 32; DB 3; Length 118;  
Best Local Similarity 75.0%; Pred. No. 1e+02;  
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 GGICVOPG 8  
11 1111  
DB 27 GGALVOPG 34

## RESULT 13

US-08-428-197-38  
Sequence 38, Application US/08428197  
Patent No. 5891438  
GENERAL INFORMATION:  
APPLICANT: SILVERMAN, GREGG J.  
TITLE OF INVENTION: METHOD FOR STIMULATING PRODUCTION OF ANTIBODIES THROUGH  
TITLE OF INVENTION: VARIABLE REGION GENE FAMILY RESTRICTED  
TITLE OF INVENTION: VACCINATION WITH A B-CELL SUPERNATIGN AND CONUGATES  
TITLE OF INVENTION: THEREOF  
NUMBER OF SEQUENCES: 51  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Spensley Horn Judas & Lubitz  
STREET: 1880 Century Park East - Suite 500  
CITY: Los Angeles  
STATE: California  
COUNTRY: USA  
ZIP: 90067  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/428,197  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US93/10555  
FILING DATE: 29-OCT-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Howells, Stacy L.  
REGISTRATION NUMBER: 34,842  
REFERENCE/DOCKET NUMBER: FD-2630  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 455-5100  
TELEFAX: (619) 455-5110  
INFORMATION FOR SEQ ID NO: 38:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 123 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-428-197-38

Query Match 68.1%; Score 32; DB 2; Length 123;  
Best Local Similarity 75.0%; Pred. No. 1.1e+02;  
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 GGICVOPG 8  
11 1111  
DB 8 GGTLVOPG 15

## RESULT 14

PCT-US93-10555-38  
Sequence 38, Application PC/TUS9310555  
GENERAL INFORMATION:  
APPLICANT: SILVERMAN, GREGG J.  
TITLE OF INVENTION: METHOD FOR STIMULATING PRODUCTION OF  
TITLE OF INVENTION: VARIABLE REGION GENE FAMILY RESTRICTED ANTIBODIES THROUGH  
TITLE OF INVENTION: VACCINATION WITH A B-CELL SUPERNATIGN AND CONUGATES  
TITLE OF INVENTION: THEREOF  
NUMBER OF SEQUENCES: 51  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Spensley Horn Judas & Lubitz  
STREET: 1880 Century Park East - Suite 500

CITY: Los Angeles  
STATE: California  
COUNTRY: USA  
ZIP: 90067

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US93/10555

FILING DATE: 29-OCT-1993

## CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:  
NAME: Howells, Stacy L.

REGISTRATION NUMBER: 34,842

REFERENCE/DOCKET NUMBER: FD-2630

TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 455-5100

TELEFAX: (619) 455-5110

INFORMATION FOR SEQ ID NO: 38:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 123 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
PCT-US93-10555-38

Query Match 68.1%; Score 32; DB 5; Length 123;  
Best Local Similarity 75.0%; Pred. No. 1.1e+02;  
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 GGICVOPG 8  
11 1111  
DB 8 GGTLVOPG 15

## RESULT 15

US-08-651-999A-1  
Sequence 1, Application US/08651999A  
Patent No. 6031088  
GENERAL INFORMATION:  
APPLICANT: Stefan Somlo and Toshio Mochizuki  
TITLE OF INVENTION: POLYCYSTIC KIDNEY DISEASE PKD2 GENE  
TITLE OF INVENTION: AND USES THEREOF  
NUMBER OF SEQUENCES: 12  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: AMSTER, ROTHSTEIN & EBENSTEIN  
STREET: 90 PARK AVENUE  
CITY: NEW YORK  
STATE: NEW YORK  
COUNTRY: U.S.A.  
ZIP: 10016  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 INCH 1.44 Mb STORAGE  
MEDIUM TYPE: DISKETTE  
COMPUTER: IBM PC COMPATIBLE  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: ASCII  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/651,999A  
FILING DATE: MAY 23, 1996  
ATTORNEY/AGENT INFORMATION:  
NAME: ELIZABETH A. BOGOSIAN  
REGISTRATION NUMBER: 39,911  
REFERENCE/DOCKET NUMBER: 96700/395  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 697-5955  
TELEFAX: (212) 286-0854 or 286-0082  
TELEX: TWX 710-581-4766  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:

LENGTH: 866 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE:  
DESCRIPTION: peptide  
HYPOTHETICAL: NO  
FRAGMENT TYPE: Internal fragment  
US-08-651-999A-1

Query Match 68.1%; Score 32; DB 3; Length 866;  
Best Local Similarity 75.0%; Pred. No. 6.8e+02;  
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GGLCVQPG 8  
      |||||  
Db 44 GGLCEORG 51

Search completed: June 13, 2001, 14:16:36  
Job time: 497 sec

*This Page Blank (uspto)*

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

# OM protein - protein search, using sw model

Run on: June 13, 2001, 14:14:39 ; Search time 229.28 Seconds  
(without alignments)  
1.995 Million cell updates/sec

Title: PCT-US01-05825A-18

Perfect score: 46  
Sequence: 1 GGLCVQDG 8

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 390729 seqs, 57163235 residues

Total number of hits satisfying chosen parameters: 390729

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_0401.\*

1:	/SID6/gcgdata/geneseq/AA1980.DAT.*
2:	/SID6/gcgdata/geneseq/AA1981.DAT.*
3:	/SID6/gcgdata/geneseq/AA1982.DAT.*
4:	/SID6/gcgdata/geneseq/AA1983.DAT.*
5:	/SID6/gcgdata/geneseq/AA1984.DAT.*
6:	/SID6/gcgdata/geneseq/AA1985.DAT.*
7:	/SID6/gcgdata/geneseq/AA1986.DAT.*
8:	/SID6/gcgdata/geneseq/AA1987.DAT.*
9:	/SID6/gcgdata/geneseq/AA1988.DAT.*
10:	/SID6/gcgdata/geneseq/AA1989.DAT.*
11:	/SID6/gcgdata/geneseq/AA1990.DAT.*
12:	/SID6/gcgdata/geneseq/AA1991.DAT.*
13:	/SID6/gcgdata/geneseq/AA1992.DAT.*
14:	/SID6/gcgdata/geneseq/AA1993.DAT.*
15:	/SID6/gcgdata/geneseq/AA1994.DAT.*
16:	/SID6/gcgdata/geneseq/AA1995.DAT.*
17:	/SID6/gcgdata/geneseq/AA1996.DAT.*
18:	/SID6/gcgdata/geneseq/AA1997.DAT.*
19:	/SID6/gcgdata/geneseq/AA1998.DAT.*
20:	/SID6/gcgdata/geneseq/AA1999.DAT.*
21:	/SID6/gcgdata/geneseq/AA2000.DAT.*
22:	/SID6/gcgdata/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	46	100.0	8	21	V79122
2	43	93.5	8	21	V79118
3	41	89.1	196	20	V36831
4	39	84.8	8	21	V79121
5	38	82.6	8	21	V79110
6	38	82.6	8	21	V79126
7	38	82.6	399	13	R20006
8	37	80.4	792	16	R85198
9	37	80.4	793	16	R85200
10	37	80.4	793	16	R85199
11	36	78.3	8	21	V79117

12	36	78.3	8	21	V79124	Peptide antagonist
13	35	76.1	8	21	V79106	Peptide antagonist
14	35	76.1	713	15	R60101	Canine zona pelluc
15	35	76.1	715	15	R5198	Canine zona pelluc
16	35	76.1	715	20	V42471	Canine zona pelluc
17	35	76.1	715	20	W81808	Canine ZPA protein
18	35	76.1	715	21	V82206	Canine zona pelluc
19	35	76.1	715	21	V52680	Canine oocyte zona
20	35	76.1	715	21	V52171	Canine zona pelluc
21	35	76.1	715	21	V52975	Canine zona pelluc
22	35	76.1	716	15	R5200	Feline zona pelluc
23	35	76.1	716	15	R60532	Feline zona pelluc
24	35	76.1	716	20	V42473	Feline zona pelluc
25	35	76.1	716	20	W81810	Feline ZPA protein
26	35	76.1	716	21	V82208	Feline zona pelluc
27	35	76.1	716	21	V52682	Feline oocyte zona
28	35	76.1	716	21	V52173	Feline zona pelluc
29	35	76.1	716	21	V52977	Feline zona pelluc
30	34	73.9	36	21	BA0159	Human secreted pro
31	34	73.9	380	18	W09406	Transforming growt
32	34	73.9	742	16	R74094	Human zona pelluci
33	34	73.9	745	15	R55206	Human zona pelluci
34	34	73.9	745	20	V42480	Human zona pelluci
35	34	73.9	745	20	W81817	Human ZPA protein
36	34	73.9	745	21	V82215	Human zona pelluci
37	34	73.9	745	21	V52689	Human oocyte zona
38	34	73.9	745	21	V52180	Human zona pelluci
39	34	73.9	745	21	V52984	Human zona pelluci
40	34	73.9	1487	19	W61562	Human type II coll
41	33	71.7	8	21	V79120	Peptide antagonist
42	33	71.7	31	20	W68384	Human zneul EGF-11
43	33	71.7	48	22	B55858	PDZ encoded domain
44	33	71.7	48	22	B57649	Human MINT3 protei
45	33	71.7	48	22	B58059	Human MINT3 protei

## ALIGNMENTS

RESULT	1
ID	V79122
Standard:	Peptide: 8 AA.
V79122:	
05-JUN-2000	(first entry)
Peptide antagonist of zonulin.	
XX	Zonulin: antagonist; zonula occludens toxin receptor;
XX	blood-brain barrier; antiinflammatory; cerebroprotective;
XX	neuroprotective; dermatological; antitumor; antiviral;
XX	antibacterial; cytostatic; anti-HIV; vulnereary; antiallergic;
XX	hypotensive; immunosuppressive; antiparasitic; vasotropic;
XX	gastrointestinal inflammation; therapy.
OS	Synthetic.
XX	
PN	W0200007609-A1.
XX	
PD	17-FEB-2000.
XX	
PF	28-JUL-1999; 99MO-US16683.
XX	
PR	03-AUG-1998; 98US-0127815.
XX	
PA	(UYMA-) UNIV MARYLAND BALTIMORE.
XX	
PI	Fasano A;
XX	
DR	WPI: 2000-205565/18.
XX	
PT	New peptide antagonist of zonulin useful as antiinflammatory agent for

PT treating cerebral ischemia, stroke, cerebral edema, gastritis,  
 PT shigellosis, viral gastroenteritis, meningitis, encephalomyelitis  
 PS  
 XX  
 Claim 1; Page 46; 69pp; English.

CC This present sequence is that of a peptide antagonist of zonulin  
 CC (2), one of 25 such peptides (see Y79105-29) of the invention,  
 CC which bind to a zonula occludens toxin (ZOT) receptor, yet do not  
 CC physiologically modulate the opening of mammalian tight junctions  
 CC (TJ). The peptide antagonists are based on a common motif of ZOT  
 CC and human zonulins, which is believed to be critical for receptor  
 CC binding. They can be prepared by chemical synthesis or by use of  
 CC recombinant DNA techniques. The peptide antagonists are used as an  
 CC antiinflammatory agents in the treatment of gastrointestinal  
 CC inflammation, where they bind to the ZOT receptor in the intestine  
 CC and yet does not physiologically modulate the opening of TJ in the  
 CC intestine. Gastrointestinal inflammation conditions give rise to  
 CC increased intestinal permeability and the peptide is useful for  
 CC treating intestinal conditions that cause protein losing enteropathy  
 CC caused by infection, e.g. Clostridium difficile infection,  
 CC enterocolitis, shigellosis, viral gastroenteritis, parasite  
 CC infestation, bacterial overgrowth, whipple's disease, diseases with  
 CC mucosal erosion or ulcerations, e.g. gastritis, gastric cancer,  
 CC collagenous colitis, inflammatory bowel disease, diseases marked by  
 CC lymphatic obstruction, e.g. congenital intestinal lymphangiectasia,  
 CC sarcoidosis lymphoma, mesenteric tuberculosis, and after surgical  
 CC correction of congenital heart disease with Fontan's operation,  
 CC mucosal diseases without ulceration, e.g. Menetrier's disease,  
 CC coliac disease, eosinophilic gastroenteritis, and immune diseases,  
 CC e.g. systemic lupus erythematosus or food allergies, primarily to  
 CC milk.  
 CC  
 XX  
 Sequence 8 AA:

Query Match 100.0%; Score 46; DB 21; Length 8;  
 Best Local Similarity 100.0%; Pred. No. 3.2e+05;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGLCVQDG 8  
 |||||  
 Db 1 gglcvqdg 8

RESULT 2  
 Y79118  
 ID Y79118 standard; Peptide; 8 AA.  
 XX  
 AC Y79118;  
 XX  
 DT 05-JUN-2000 (first entry)  
 XX  
 DE Peptide antagonist of zonulin.

XX Zonulin; antagonist; zonula occludens toxin receptor;  
 KM blood-brain barrier; antiinflammatory; cerebroprotective;  
 KM neuroprotective; dermatological; antiulcer; antiviral;  
 KM antibacterial; cytostatic; anti-HIV; vulnereary; antiallergic;  
 KM hypotensive; immunosuppressive; antiparasitic; vasotropic;  
 KM gastrointestinal inflammation; therapy.  
 XX  
 OS Synthetic.  
 XX  
 PN WO200007609-A1.  
 XX  
 PD 17-FEB-2000.  
 XX  
 PF 28-JUL-1999; 99MO-US16683.  
 XX  
 PR 03-AUG-1998; 98US-0127815.  
 XX  
 PA (UYMA-) UNIV MARYLAND BALTIMORE.  
 XX

PI Fasano A;  
 XX  
 DR WPI; 2000-205565/18.  
 XX  
 PT  
 PT New peptide antagonist of zonulin useful as antiinflammatory agent for  
 PT treating cerebral ischemia, stroke, cerebral edema, gastritis,  
 PT shigellosis, viral gastroenteritis, meningitis, encephalomyelitis  
 PS  
 XX  
 Claim 1; Page 45; 69pp; English.

CC This present sequence is that of a peptide antagonist of zonulin  
 CC (2), one of 25 such peptides (see Y79105-29) of the invention,  
 CC which bind to a zonula occludens toxin (ZOT) receptor, yet do not  
 CC physiologically modulate the opening of mammalian tight junctions  
 CC (TJ). The peptide antagonists are based on a common motif of ZOT  
 CC and human zonulins, which is believed to be critical for receptor  
 CC binding. They can be prepared by chemical synthesis or by use of  
 CC recombinant DNA techniques. The peptide antagonists are used as an  
 CC antiinflammatory agents in the treatment of gastrointestinal  
 CC inflammation, where they bind to the ZOT receptor in the intestine  
 CC and yet does not physiologically modulate the opening of TJ in the  
 CC intestine. Gastrointestinal inflammation conditions give rise to  
 CC increased intestinal permeability and the peptide is useful for  
 CC treating intestinal conditions that cause protein losing enteropathy  
 CC caused by infection, e.g. Clostridium difficile infection,  
 CC enterocolitis, shigellosis, viral gastroenteritis, parasite  
 CC infestation, bacterial overgrowth, whipple's disease, diseases with  
 CC mucosal erosion or ulcerations, e.g. gastritis, gastric cancer,  
 CC collagenous colitis, inflammatory bowel disease, diseases marked by  
 CC lymphatic obstruction, e.g. congenital intestinal lymphangiectasia,  
 CC sarcoidosis lymphoma, mesenteric tuberculosis, and after surgical  
 CC correction of congenital heart disease with Fontan's operation,  
 CC mucosal diseases without ulceration, e.g. Menetrier's disease,  
 CC coliac disease, eosinophilic gastroenteritis, and immune diseases,  
 CC e.g. systemic lupus erythematosus or food allergies, primarily to  
 CC milk.  
 CC  
 XX  
 Sequence 8 AA:

Query Match 93.5%; Score 43; DB 21; Length 8;  
 Best Local Similarity 87.5%; Pred. No. 3.2e+05;  
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGLCVQDG 8  
 |||||  
 Db 1 gglcvqdg 8

RESULT 3  
 Y36831  
 ID Y36831 standard; Protein; 196 AA.  
 XX  
 AC Y36831;  
 XX  
 DT 07-OCT-1999 (first entry)  
 XX  
 DE Amino acid sequence of a Chlamydia trachomatis protein.

XX Vaccine; eye disease; conventional trachoma; nonendemic trachoma;  
 KM paratrachoma; inclusion conjunctivitis; genital disease; periteneitis;  
 KM nongonococcal urethritis; epididymitis; cervicitis; salpingitis;  
 KM bartolinitis; pneumonia; venereal lymphogranulomatosis.  
 XX  
 OS Chlamydia trachomatis.  
 XX  
 PN WO9928475-A2.  
 XX  
 PD 10-JUN-1999.  
 XX  
 PF 27-NOV-1998; 98MO-IB01939.  
 XX  
 PR 04-NOV-1998; 98US-0107077.



PF	28-NOV-1997:	97ER-0015041.
PR	17-DEC-1997:	97ER-0016034.
XX	(GEST ) GENSET.	
PA	Griffais R;	
XX		
PI	WPI: 1999-371125/31.	
DR		
XX	Genome sequence of Chlamydia trachomatis	
XX	Disclosure: Page 710-711; 1755pp; English.	
PS		
CC	Y36754-Y37949 are encoded by open reading frames (ORFs) of the genome of	
CC	Chlamydia trachomatis (see 201425). The polypeptides can be used as	
CC	vaccines against Chlamydia trachomatis. Antisense and ribozyme sequences	
CC	can also be used to control growth of the microorganism. Chlamydia	
CC	trachomatis is responsible for a large number of diseases, e.g. eye	
CC	diseases such as conventional trachoma, nongonococcal trachoma,	
CC	paratrachoma, and inclusion conjunctivitis; genital diseases such as	
CC	nongonococcal urethritis, epididymitis, cervicitis, salpingitis,	
CC	perithelitis, Bartholinitis; pneumonia in breast feeding infants;	
CC	and venereal lymphogranulomatosis. The polypeptides of the invention	
CC	may be of use in treating these diseases.	
SO	Sequence 196 AA:	
XX		
XX	Query Match	89.1%; Score 41; DB 20; Length 196;
XX	Best Local Similarity	87.5%; Pred. NO. 10;
XX	Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0.	
QY	1 GGLCYQDG 8	
	111111	
DB	150 ggycvvvg 157	
RESULT 4		
ID	Y79121 standard; Peptide; 8 AA.	
XX	Y79121;	
XX	05-JUN-2000 (first entry)	
DT		
XX	Peptide antagonist of zonulin.	
DE		
KM	Zonulin: antagonist; zonula occludens toxin receptor;	
KM	blood-brain barrier; antiinflammatory; cerebroprotective;	
KM	neuroprotective; dermatological; antitumor; antiviral;	
KM	antibacterial; cytostatic; anti-HIV; vulvular; antiallergic;	
KM	hypotensive; immunosuppressive; antiparasitic; vasotropic;	
KM	gastrointestinal inflammation; therapy.	
XX		
XX	Synthetic.	
OS		
XX	WO200007609-A1.	
XX	17-FEB-2000.	
PD		
XX	28-JUL-1999; 99WO-US16683.	
PF		
XX	03-AUG-1998; 98US-0127815.	
PR		
XX	(UYMA-) UNIV MARYLAND BALTIMORE.	
PA		
XX	Pasano A;	
PI		
XX	WPI: 2000-205565/18.	
DR		
XX	New peptide antagonist of zonulin useful as antiinflammatory agent for	
PT	treating cerebral ischemia, stroke, cerebral edema, gastritis,	
PT	shigellosis, viral gastroenteritis, meningitis, encephalomyelitis -	

Claim 1; Page 46; 69pp; English.  
 This present sequence is that of a peptide antagonist of zonulin (2), one of 25 such peptides (see Y79105-29) of the invention, which bind to a zonula occludens toxin (ZOT) receptor, yet do not physiologically modulate the opening of mammalian tight junctions (TJ). The peptide antagonists are based on a common motif of ZOT and human zonulins, which is believed to be critical for receptor binding. They can be prepared by chemical synthesis or by use of recombinant DNA techniques. The peptide antagonists are used as an antiinflammatory agents in the treatment of gastrointestinal inflammation, where they bind to the ZOT receptor in the intestine and yet does not physiologically modulate the opening of TJ in the intestine. Gastrointestinal inflammation conditions give rise to increased intestinal permeability and the peptide is useful for treating intestinal conditions that cause protein losing enteropathy caused by infection, e.g. Clostridium difficile infection, enterocolitis, shigellosis, viral gastroenteritis, parasite infestation, bacterial overgrowth, whipple's disease, diseases with mucosal erosion or ulcerations, e.g. gastritis, gastric cancer, collagenous colitis, inflammatory bowel disease, diseases marked by lymphatic obstruction, e.g. congenital intestinal lymphangiectasia, sarcoidosis lymphoma, mesenteric tuberculosis, and after surgical correction of congenital heart disease with Fontan's operation, mucosal diseases without ulceration, e.g. Menetrier's disease, Crohn's disease, eosinophilic gastroenteritis, and immune diseases, e.g. systemic lupus erythematosus or food allergies, primarily to milk.  
 Sequence 8 AA:  
 Query Match 84.8%; Score 39; DB 21; Length 8;  
 Best Local Similarity 87.5%; Pred. No. 3.2e+05;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0.  
 Oy 1 GGLCVQDG 8  
 |||||  
 Db 1 gglcvqpg 8  
 RESULT 5  
 Y79110  
 ID Y79110 standard; Peptide: 8 AA.  
 AC Y79110:  
 XX 05-JUN-2000 (first entry)  
 DT  
 DE Peptide antagonist of zonulin.  
 KW Zonulin: antagonist; zonula occludens toxin receptor;  
 KW blood-brain barrier; antiinflammatory; cerebroprotective;  
 KW neuroprotective; dermatological; antitumor; antiviral;  
 KW antibacterial; cytostatic; anti-HIV; vulnery; antiallergic;  
 KW hypotensive; immunosuppressive; antiparasitic; vasotropic;  
 KW gastrointestinal inflammation; therapy.  
 OS Synthetic.  
 PN WO200007609-A1.  
 PD 17-FEB-2000.  
 PF 28-JUL-1999; 99WO-US16683.  
 PR 03-AUG-1998; 98US-0127815.  
 RA (UYMA-) UNIV MARYLAND BALTIMORE.  
 RI Fasano A;  
 XX

DR WPI; 2000-205565/18.

XX New peptide antagonist of zonulin useful as antiinflammatory agent for  
PT treating cerebral ischemia, stroke, cerebral edema, gastritis,  
XX shigellosis, viral gastroenteritis, meningitis, encephalomyelitis  
PS Claim 1; Page 42; 69pp; English.

CC This present sequence is that of a peptide antagonist of zonulin  
CC (2), one of 25 such peptides (see Y79105-29) of the invention,  
CC which bind to a zonula occludens toxin (ZOT) receptor, yet do not  
CC physiologically modulate the opening of mammalian tight junctions  
CC (TJ). The peptide antagonists are based on a common motif of ZOT  
CC and human zonulins, which is believed to be critical for receptor  
CC binding. They can be prepared by chemical synthesis or by use of  
CC recombinant DNA techniques. The peptide antagonists are used as an  
CC antiinflammatory agents in the treatment of gastrointestinal  
CC inflammation, where they bind to the ZOT receptor in the intestine  
CC and yet does not physiologically modulate the opening of TJ in the  
CC intestine. Gastrointestinal inflammation conditions give rise to  
CC increased intestinal permeability and the peptide is useful for  
CC treating intestinal conditions that cause protein losing enteropathy  
CC caused by infection, e.g. Clostridium difficile infection,  
CC enterocolitis, shigellosis, viral gastroenteritis, parasite  
CC infestation, bacterial overgrowth, whipple's disease, diseases with  
CC mucosal erosion or ulcerations, e.g. gastritis, gastric cancer,  
CC collagenous colitis, inflammatory bowel disease, diseases marked by  
CC lymphatic obstruction, e.g. congenital intestinal lymphangiectasia,  
CC sarcoidosis lymphoma, mesenteric tuberculosis, and after surgical  
CC correction of congenital heart disease with Fontan's operation,  
CC mucosal diseases without ulceration, e.g. Menetrier's disease,  
CC coeliac disease, eosinophilic gastroenteritis, and immune diseases,  
CC e.g. systemic lupus erythematosus or food allergies, primarily to  
CC milk.

SO Sequence 8 AA;

Query Match 82.6%; Score 38; DB 21; Length 8;  
Best Local Similarity 87.5%; Pred. No. 3.2e+05;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GGLCVODG 8  
| | | | | | | |  
DB 1 grlcvqdg 8

RESULT 6  
Y79126  
ID Y79126 standard; Peptide: 8 AA.

XX Y79126;

DT 05-JUN-2000 (first entry)

DE Peptide antagonist of zonulin.  
XX  
XX Zonulin; antagonist; zonula occludens toxin receptor;  
KM blood-brain barrier; antiinflammatory; cerebroprotective;  
KM neuroprotective; dermatological; antitumor; antiviral;  
KM antibacterial; cytostatic; anti-HIV; vulnereary; antiallergic;  
KM hypotensive; immunosuppressive; antiparasitic; vasotropic;  
KM gastrointestinal inflammation; therapy.

XX Synthetic.

XX MO200007609-A1.

PD 17-FEB-2000.

PF 28-JUL-1999; 99WO-US16683.

XX 03-AUG-1998; 98US-0127815.

XX (UYMA-) UNIV MARYLAND BALTIMORE.

XX Fasanio A;

DR WPI; 2000-205565/18.

XX New peptide antagonist of zonulin useful as antiinflammatory agent for  
PT treating cerebral ischemia, stroke, cerebral edema, gastritis,  
XX shigellosis, viral gastroenteritis, meningitis, encephalomyelitis  
PS Claim 1; Page 48; 69pp; English.

CC This present sequence is that of a peptide antagonist of zonulin  
CC (2), one of 25 such peptides (see Y79105-29) of the invention,  
CC which bind to a zonula occludens toxin (ZOT) receptor, yet do not  
CC physiologically modulate the opening of mammalian tight junctions  
CC (TJ). The peptide antagonists are based on a common motif of ZOT  
CC and human zonulins, which is believed to be critical for receptor  
CC binding. They can be prepared by chemical synthesis or by use of  
CC recombinant DNA techniques. The peptide antagonists are used as an  
CC antiinflammatory agents in the treatment of gastrointestinal  
CC inflammation, where they bind to the ZOT receptor in the intestine  
CC and yet does not physiologically modulate the opening of TJ in the  
CC intestine. Gastrointestinal inflammation conditions give rise to  
CC increased intestinal permeability and the peptide is useful for  
CC treating intestinal conditions that cause protein losing enteropathy  
CC caused by infection, e.g. Clostridium difficile infection,  
CC enterocolitis, shigellosis, viral gastroenteritis, parasite  
CC infestation, bacterial overgrowth, whipple's disease, diseases with  
CC mucosal erosion or ulcerations, e.g. gastritis, gastric cancer,  
CC collagenous colitis, inflammatory bowel disease, diseases marked by  
CC lymphatic obstruction, e.g. congenital intestinal lymphangiectasia,  
CC sarcoidosis lymphoma, mesenteric tuberculosis, and after surgical  
CC correction of congenital heart disease with Fontan's operation,  
CC mucosal diseases without ulceration, e.g. Menetrier's disease,  
CC coeliac disease, eosinophilic gastroenteritis, and immune diseases,  
CC e.g. systemic lupus erythematosus or food allergies, primarily to  
CC milk.

SO Sequence 8 AA;

Query Match 82.6%; Score 38; DB 21; Length 8;  
Best Local Similarity 87.5%; Pred. No. 3.2e+05;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GGLCVODG 8  
| | | | | | | |  
DB 1 gggcvqdg 8

RESULT 7  
R20006  
ID R20006 standard; Protein: 399 AA.

XX R20006;

DT 31-MAR-1992 (first entry)

DE Zonula occludens toxin.  
XX  
XX ZOT; cholera; vaccine; enterotoxin; diarrhoea.

XX Vibrio cholerae.

XX MO9118979-A.

PD 12-DEC-1991.

PF 05-JUN-1991; 91WO-US03812.

XX 05-JUN-1990; 90US-0533315.

XX (UYMA-) UNIV MARYLAND BALTI.  
 PA Kaper JB, Baudry-Maurelli B, Fasano A;  
 XX MPI; 1992-007465/01.  
 DR N-PSDB; Q20185.  
 XX  
 PT New Vibrio cholerae strains - comprise restriction endonuclease  
 PT fragment encoding toxin, used as vaccines against cholera  
 XX  
 PS Disclosure: Fig 18; 83pp; English.  
 XX  
 CC The amino acid sequence is that of the zonula occludens toxin (ZOT).  
 CC It may be responsible for diarrhoea in some strains of cholera and  
 CC the ZOT gene or fragments of it are deleted from strains of Vibrio  
 CC cholerae (V.c.) to be used as vaccines. These V.c. strains have 100%  
 CC efficiency in protecting humans against subsequent infection with a  
 CC strain of a similar serotype and avoid undesirable side effects such  
 CC as diarrhoea, nausea and cramping. Cultures of these strains may be  
 CC used for prodn. of vaccines against cholera.  
 CC  
 SQ Sequence 399 AA;  
 XX  
 Query Match 82.6%; Score 38; DB 13; Length 399;  
 Best Local Similarity 87.5%; Pred. NO. 62;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 GGLCVQDG 8  
 1 111111  
 Db 291 grlcvgdg 298  
 XX  
 RESULT 8  
 R85198  
 ID R85198 standard; Protein; 792 AA.  
 XX  
 AC R85198;  
 XX  
 DT 25-JUN-1996 (first entry)  
 XX  
 DE Avenacinase - a saponin glycosyl hydrolase.  
 XX  
 KW saponin glycosyl hydrolase; tomatinase; plant pathogenic fungi;  
 KW avenacinase; deglycosylation; pore formation; cell death.  
 XX  
 OS Gaemannomyces graminis var. avenae.  
 OS  
 PN W09530009-A2.  
 XX  
 PD 09-NOV-1995.  
 XX  
 PE 17-MAR-1995; 95WO-GB00592.  
 XX  
 PR 29-APR-1994; 94GB-0008573.  
 XX  
 PA (GATS-) GATSBY CHARITABLE FOUND.  
 PA  
 PI Bowyer P, Daniels MJ, Osbourn AE;  
 XX  
 DR MPI; 1995-393080/50.  
 DR N-PSDB; T06022.  
 XX  
 PT New isolated saponin glycosyl hydrolase enzymes - used to develop  
 PT prods. for the modification of microbial organisms and plants or  
 PT plant prods.  
 XX  
 PS Claim 2; Fig 4c; 113pp; English.  
 PS  
 CC Avenacinase is a saponin glycosyl hydrolase (SGH) encoded by T06022,  
 CC isolated from the plant pathogenic fungi Gaemannomyces graminis var.  
 CC avenae. Avenacinase detoxifies the triterpenoid saponin avenacin A-1

CC found in oat roots, by hydrolytic cleavage of the beta, 1-2 and beta,  
 CC 1-4 linked terminal glucose mol. to give sequentially the mono-  
 CC and bis-deglucosylated forms, both of which are substantially less toxic  
 CC to fungal growth than avenacin A-1 itself. This deglycosylation is  
 CC sufficient to destroy the ability of the saponin to complex with membrane  
 CC sterols. Saponin/sterol complexes in eukaryotic membranes results in pore  
 CC formation and leakage of cell contents, with subsequent cell death. The  
 CC DNA and proteins of the invention are useful in identification of related  
 CC enzymes, structural studies of saponins and also for development of  
 CC agents which can modulate SGH activity, e.g. for reducing pathogenicity  
 CC of SGH-producing pathogens for specific hosts.  
 CC  
 SQ Sequence 792 AA;  
 XX  
 Query Match 80.4%; Score 37; DB 16; Length 792;  
 Best Local Similarity 85.7%; Pred. NO. 1.7e+02;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 2 GGLCVQDG 8  
 111111  
 Db 88 glicldqg 94  
 XX  
 RESULT 9  
 R85200  
 ID R85200 standard; Protein; 793 AA.  
 XX  
 AC R85200;  
 XX  
 DT 04-JUL-1996 (first entry)  
 XX  
 DE Avenacinase-like protein.  
 XX  
 KW saponin glycosyl hydrolase; tomatinase; plant pathogenic fungi;  
 KW avenacinase; deglycosylation; pore formation; cell death; ss.  
 XX  
 OS Gaemannomyces graminis var. tritici.  
 OS  
 PN W09530009-A2.  
 XX  
 PD 09-NOV-1995.  
 XX  
 PE 17-MAR-1995; 95WO-GB00592.  
 XX  
 PR 29-APR-1994; 94GB-0008573.  
 XX  
 PA (GATS-) GATSBY CHARITABLE FOUND.  
 PA  
 PI Bowyer P, Daniels MJ, Osbourn AE;  
 XX  
 DR MPI; 1995-393080/50.  
 DR N-PSDB; T06024.  
 XX  
 PT New isolated saponin glycosyl hydrolase enzymes - used to develop  
 PT prods. for the modification of microbial organisms and plants or  
 PT plant prods.  
 XX  
 PS Claim 2; Fig 4e; 113pp; English.  
 PS  
 CC The saponin glycosyl hydrolase (SGH) designated avenacinase-like protein  
 CC is encoded by T06024, and was isolated from the plant pathogenic fungi  
 CC Gaemannomyces graminis var. tritici. Avenacinase detoxifies the  
 CC triterpenoid saponin avenacin A-1 found in oat roots, by hydrolytic  
 CC cleavage of the beta, 1-2 and beta, 1-4 linked terminal glucose mol. to  
 CC give sequentially the mono- and bis-deglucosylated forms, both of which  
 CC are substantially less toxic to fungal growth than avenacin A-1 itself.  
 CC This deglycosylation is sufficient to destroy the ability of the saponin  
 CC to complex with membrane sterols. Saponin/sterol complexes in eukaryotic  
 CC membranes results in pore formation and leakage of cell contents, with  
 CC subsequent cell death. The DNA and proteins of the invention are useful  
 CC in identification of related enzymes, structural studies of saponins and  
 CC also for development of agents which can modulate SGH activity, e.g. for

CC lymphatic obstruction, e.g. congenital lymphangioectasia

CC correction of congenital heart disease with Fontan's operation,  
CC mucosal diseases without ulceration, e.g. Menetrier's disease,  
CC coeliac disease, eosinophilic gastroenteritis, and immune diseases,  
CC e.g. systemic lupus erythematosus or food allergies, primarily to  
CC milk.  
XX  
SO Sequence 8 AA:  
  
Query Match 78.3%; Score 36; DB 21; Length 8;  
Best Local Similarity 75.0%; Pred. No. 3.2e+05;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
  
QY 1 GGLCYQDG 8  
Db 1 ggvccvpg 8  
11:11111  
1 ggvccvpg 8  
  
RESULT 12  
Y79124  
ID Y79124 standard; Peptide: 8 AA.  
XX Y79124;  
AC  
XX Y79124;  
DE 05-JUN-2000 (first entry)  
XX  
DE Peptide antagonist of zonulin.  
XX  
KW Zonulin; antagonist; zonula occludens toxin receptor;  
KW blood-brain barrier; antiinflammatory; cerebroprotective;  
KW neuroprotective; dermatological; antiulcer; antiviral;  
KW antibacterial; cytostatic; anti-HIV; vulnery; antiallergic;  
KW hypotensive; immunosuppressive; antiparasitic; vasotropic;  
KW gastrointestinal inflammation; therapy.  
XX  
OS Synthetic.  
XX  
PN WO200007609-A1.  
XX 17-FEB-2000.  
PD  
XX  
XX 28-JUL-1999; 99WO-US16683.  
PF  
XX 03-AUG-1998; 98US-0127815.  
PR  
XX (UYMA-) UNIV MARYLAND BALTIMORE.  
PA  
XX Fasano A;  
PI  
XX WPI: 2000-205565/18.  
DR  
XX  
XX New peptide antagonist of zonulin useful as antiinflammatory agent for  
PT treating cerebral ischemia, stroke, cerebral edema, gastritis,  
PT shigellosis, viral gastroenteritis, meningitis, encephalomyelitis  
PS  
XX Claim 1; Page 47; 69pp; English.  
XX  
XX This present sequence is that of a peptide antagonist of zonulin  
CC (Z), one of 25 such peptides (see Y79105-29) of the invention,  
CC which bind to a zonula occludens toxin (ZOT) receptor, yet do not  
CC physiologically modulate the opening of mammalian tight junctions  
CC (TJ). The peptide antagonists are based on a common motif of ZOT  
CC and human zonulins, which is believed to be critical for receptor  
CC binding. They can be prepared by chemical synthesis or by use of  
CC recombinant DNA techniques. The peptide antagonists are used as an  
CC antiinflammatory agents in the treatment of gastrointestinal  
CC inflammation, where they bind to the ZOT receptor in the intestine  
CC and yet does not physiologically modulate the opening of TJ in the  
CC intestine. Gastrointestinal inflammation conditions give rise to  
CC increased intestinal permeability and the peptide is useful for  
CC treating intestinal conditions that cause protein losing enteropathy  
CC caused by infection, e.g. Clostridium difficile infection,  
CC enterocolitis, shigellosis, viral gastroenteritis, parasite

CC infestation, bacterial overgrowth, whipple's disease, diseases with  
CC mucosal erosion or ulcerations, e.g. gastritis, gastric cancer,  
CC collagenous colitis, inflammatory bowel disease, diseases marked by  
CC lymphatic obstruction, e.g. congenital intestinal lymphangiectasia,  
CC sarcoidosis lymphoma, mesenteric tuberculosis, and after surgical  
CC correction of congenital heart disease with Fontan's operation,  
CC mucosal diseases without ulceration, e.g. Menetrier's disease,  
CC coeliac disease, eosinophilic gastroenteritis, and immune diseases,  
CC e.g. systemic lupus erythematosus or food allergies, primarily to  
CC milk.  
XX  
SO Sequence 8 AA:  
  
Query Match 78.3%; Score 36; DB 21; Length 8;  
Best Local Similarity 87.5%; Pred. No. 3.2e+05;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
  
QY 1 GGLCYQDG 8  
Db 1 ggllyvqdg 8  
11111111  
1 ggllyvqdg 8  
  
RESULT 13  
Y79106  
ID Y79106 standard; Peptide: 8 AA.  
XX Y79106;  
AC  
XX Y79106;  
DE 05-JUN-2000 (first entry)  
XX  
DE Peptide antagonist of zonulin.  
XX  
KW Zonulin; antagonist; zonula occludens toxin receptor;  
KW blood-brain barrier; antiinflammatory; cerebroprotective;  
KW neuroprotective; dermatological; antiulcer; antiviral;  
KW antibacterial; cytostatic; anti-HIV; vulnery; antiallergic;  
KW hypotensive; immunosuppressive; antiparasitic; vasotropic;  
KW gastrointestinal inflammation; therapy.  
XX  
OS Synthetic.  
XX  
PN WO200007609-A1.  
XX 17-FEB-2000.  
PD  
XX  
XX 28-JUL-1999; 99WO-US16683.  
PF  
XX 03-AUG-1998; 98US-0127815.  
PR  
XX (UYMA-) UNIV MARYLAND BALTIMORE.  
PA  
XX Fasano A;  
PI  
XX WPI: 2000-205565/18.  
DR  
XX  
XX New peptide antagonist of zonulin useful as antiinflammatory agent for  
PT treating cerebral ischemia, stroke, cerebral edema, gastritis,  
PT shigellosis, viral gastroenteritis, meningitis, encephalomyelitis  
PS  
XX Claim 1; Page 41; 69pp; English.  
XX  
XX This present sequence is that of a peptide antagonist of zonulin  
CC (Z), one of 25 such peptides (see Y79105-29) of the invention,  
CC which bind to a zonula occludens toxin (ZOT) receptor, yet do not  
CC physiologically modulate the opening of mammalian tight junctions  
CC (TJ). The peptide antagonists are based on a common motif of ZOT  
CC and human zonulins, which is believed to be critical for receptor  
CC binding. They can be prepared by chemical synthesis or by use of  
CC recombinant DNA techniques. The peptide antagonists are used as an  
CC antiinflammatory agents in the treatment of gastrointestinal  
CC inflammation, where they bind to the ZOT receptor in the intestine  
CC and yet does not physiologically modulate the opening of TJ in the

CC Intestine. Gastrointestinal inflammation conditions give rise to  
 CC increased intestinal permeability and the peptide is useful for  
 CC treating intestinal conditions that cause protein losing enteropathy  
 CC caused by infection, e.g. Clostridium difficile infection,  
 CC enterocolitis, shigellosis, viral gastroenteritis, parasite  
 CC infestation, bacterial overgrowth, whipple's disease, diseases with  
 CC mucosal erosion or ulcerations, e.g. gastritis, gastric cancer,  
 CC collagenous colitis, inflammatory bowel disease, diseases marked by  
 CC lymphatic obstruction, e.g. congenital intestinal lymphangiectasia,  
 CC sarcoidosis lymphoma, mesenteric tuberculosis, and after surgical  
 CC correction of congenital heart disease with Fontan's operation,  
 CC mucosal diseases without ulceration, e.g. Menetrier's disease,  
 CC coeliac disease, eosinophilic gastroenteritis, and immune diseases,  
 CC e.g. systemic lupus erythematosus or food allergies, primarily to  
 CC milk.  
 CC  
 SQ Sequence 8 AA;

Query Match 76.1%; Score 35; DB 21; Length 8;  
 Best Local Similarity 75.0%; Pred. No. 3.2e+05;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 GGLCYODG 8  
 1 :|||||  
 Db 1 grvcvqdg 8

RESULT 14  
 R60101  
 ID R60101 standard: Protein; 713 AA.  
 AC R60101;  
 XX  
 DT 15-MAR-1995 (first entry)  
 XX  
 DE Canine zona pellucida CZP2.  
 KW Canine; dog; zona pellucida; ZP; CZP2; contraceptive; vaccine;  
 XX antigen.  
 XX  
 OS Canis familiaris.  
 XX  
 PN JP06189766-A.  
 PD 12-JUL-1994.  
 XX  
 PF 25-DEC-1992: 92JP-0359265.  
 XX  
 PR 25-DEC-1992: 92JP-0359265.  
 XX  
 PA (TOFU) TONEN CORP.  
 XX  
 DR WPI: 1994-259553/32.  
 DR N-PSDB: Q70072.  
 XX  
 PT New DNA sequence encoding canine zona pellucida CZP2 - useful for  
 XX the prodn. of a canine contraceptive vaccine antigen  
 XX  
 PS Claim 1; Page 8-10; 10pp; Japanese.  
 XX  
 CC The CZP2 DNA (Q70072) was prep'd. by the cloning of CZP2(75-520) -  
 CC Q81700 using the primers given in Q70073-74, CZP2(1-65) - Q81804  
 CC using the primers given in Q70082-83, CZP2(42-103) - Q81803 using  
 CC the primers given in Q70079-81 and CZP2(487-713) - Q81957 using the  
 CC primers given in Q70075-78.  
 CC  
 SQ Sequence 713 AA;

Query Match 76.1%; Score 35; DB 15; Length 713;  
 Best Local Similarity 75.0%; Pred. No. 3.4e+02;  
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 GGLCYODG 8  
 1 :|||||  
 Db 364 gdlctqdg 371

RESULT 15  
 R55198  
 ID R55198 standard: Protein; 715 AA.  
 AC R55198;  
 XX  
 DT 31-JAN-1995 (first entry)  
 XX  
 DE Canine zona pellucida ZPA protein.  
 XX  
 KW Dog; canine; zona pellucida; ZPA; immunocontraception.  
 XX  
 OS Canis familiaris.  
 XX  
 FH Key Location/Qualifiers  
 FT Protein 1..715  
 FT /label= canine\_ZPA  
 XX  
 PN W09411019-A.  
 PD 26-MAY-1994.  
 XX  
 PF 06-NOV-1993: 93WO-US10851.  
 XX  
 PR 09-NOV-1992: 92US-0973341.  
 PR 29-JAN-1993: 93US-0012390.  
 XX  
 PA (ZONA-) ZONAGEN INC.  
 XX  
 PI Harris JD, Hsu KT, Podolski JS;  
 XX  
 DR WPI: 1994-183156/22.  
 DR N-PSDB: Q65608.  
 XX  
 PT Use of zona pellucida proteins and antibodies - for inducing  
 PT reproducible transient infertility or permanent sterility in  
 PT female mammals.  
 XX  
 PS Claim 40; Page 88-90; 154pp; English.  
 XX  
 CC A commercially available 16 week old canine ovarian cDNA expression  
 CC library in lambda gt10 was screened using antibodies raised against  
 CC heat solubilised canine zona pellucida. The largest candidate clone  
 CC was used to rescreen the library and to isolate clones which were  
 CC used as probes in Southern hybridisations. Sequences coding for  
 CC canine ZPA and ZPC proteins were obtained (065608 and 065609,  
 CC respectively). R55198 is the deduced amino acid sequence for ZPA.  
 CC  
 SQ Sequence 715 AA;

Query Match 76.1%; Score 35; DB 15; Length 715;  
 Best Local Similarity 75.0%; Pred. No. 3.4e+02;  
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 GGLCYODG 8  
 1 :|||||  
 Db 366 gdlctqdg 373

Search completed: June 13, 2001, 14:14:39  
 Job time: 380 sec



*This Page Blank (uspto)*



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

## OM protein - protein search, using sw model

Run on: June 13, 2001, 14:10:47 ; Search time 130.61 Seconds

(without alignments)  
4.209 Million cell updates/sec

Title: PCT-US01-05825A-18

Perfect score: 46  
Sequence: 1 GGLCVQDG 8

## Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 198801 seqs, 68722935 residues

Total number of hits satisfying chosen parameters: 198801

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

## Database :

1: PIR-67:\*  
2: PIR1:\*  
3: PIR3:\*  
4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	41	89.1	196	F71525	hypothetical prote
2	41	89.1	196	H81681	maf protein TC0628
3	38	82.6	399	B82197	zona occludens tox
4	38	82.6	399	A43864	zonula occludens t
5	37	80.4	861	A48825	Notch homolog Motc
6	37	80.4	2531	A46019	Notch-1 protein -
7	36	78.3	633	T47524	hypothetical prote
8	36	78.3	788	Q0BEC3	HHLF1 protein - hu
9	36	78.3	846	Q0BEC3	HHLF1 protein - hu
10	36	78.3	1677	T14267	HOXFI protein, stage
11	35	76.1	296	T26723	hypothetical prote
12	35	76.1	491	D83353	mannitol dehydroge
13	35	76.1	680	T30620	hypothetical prote
14	35	76.1	715	S70397	hypothetical prote
15	35	76.1	716	S70398	zona pellucida gly
16	35	76.1	860	JC4939	beta-glucosidase (
17	34	73.9	153	A27179	collagen alpha 1(I
18	34	73.9	166	F70797	hypothetical prote
19	34	73.9	233	F75213	hypothetical prote
20	34	73.9	340	S69194	M4-(beta-N-acetyl
21	34	73.9	355	S62565	transmembrane III
22	34	73.9	380	G01639	transmembrane prot
23	34	73.9	444	1	tryptophan 5-mono
24	34	73.9	444	1	tryptophan 5-mono
25	34	73.9	444	1	tryptophan 5-mono
26	34	73.9	447	A34582	tryptophan 5-mono
27	34	73.9	520	F83157	hypothetical prote
28	34	73.9	676	2	sperm-binding gly
29	34	73.9	745	1	sperm-binding gly

30	34	73.9	1031	2	T06130	hypothetical prote
31	34	73.9	1487	1	CGHUC6	collagen alpha 1(I
32	34	73.9	1492	2	A40333	collagen alpha 1(I
33	34	73.9	1548	2	S34583	serine proteinase
34	34	73.9	2108	2	H70819	probable polyketid
35	33.5	72.8	704	2	A48040	meprin A (EC 3.4.2
36	33	71.7	115	1	A46279	guanylin precursor
37	33	71.7	142	2	S50662	hypothetical prote
38	33	71.7	161	2	S61389	small basic protei
39	33	71.7	172	2	S27022	fibroblast growth
40	33	71.7	243	2	T34150	hypothetical prote
41	33	71.7	256	2	T16805	hypothetical prote
42	33	71.7	266	2	S71025	lipopolysaccharide
43	33	71.7	274	2	F75161	homoserine kinase
44	33	71.7	362	2	S37220	fiber protein - hu
45	33	71.7	362	2	S40092	fiber protein - hu

## ALIGNMENTS

RESULT 1  
F71525  
hypothetical protein CT349 - Chlamydia trachomatis (serotype D, strain UW3/Cx)  
C:Species: Chlamydia trachomatis  
C:Date: 13-Sep-1998 #sequence\_revision 13-Sep-1998 #text\_change 08-Oct-1999  
R:Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitche  
Science 282, 754-759, 1998  
A:Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia t  
A:Reference number: A71570; MUID:99000809  
A:Accession: F71525  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-196 <ARN>  
A:Cross-references: GB:AE001308; GB:AE001273; NID:g3328766; PIDN:AAC67944.1; PID:g332  
A:Experimental source: serotype D, strain UW-3/Cx  
C:Genetics:  
A:Gene: CT349  
C:Superfamily: septum formation protein maf

Query Match 89.1%; Score 41; DB 2; Length 196;  
Best Local Similarity 87.5%; Pred. No. 1.2;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGLCVQDG 8  
DB 150 GGLCVQDG 157

RESULT 2  
H81681  
maf protein TC0628 [imported] - Chlamydia muridarum (strain N199)  
C:Species: Chlamydia muridarum, Chlamydia trachomatis MOPN  
C:Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 19-May-2000  
C:Accession: H81681  
R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hicke  
C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzbe  
Nucleic Acids Res. 28, 1397-1406, 2000  
A:Title: Genome sequences of Chlamydia trachomatis MOPN and Chlamydia pneumoniae AR39  
A:Reference number: AB1500; MUID:20150255  
A:Accession: H81681  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-196 <TEP>  
A:Cross-references: GB:AE002331; GB:AE002160; NID:g7190662; PIDN:AAF39457.1; PID:g719  
A:Experimental source: strain N199 (MOPN)  
C:Genetics:  
A:Gene: TC0628  
C:Superfamily: septum formation protein maf

```

Query Match          89.1%; Score 41; DB 2; Length 196;
Best Local Similarity 87.5%; Pred. No. 1.2;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGLCYVDG 8
    1 111111
DB 150 GGYCVODG 157

RESULT 3
B82197
zona occludens toxin VC1458 [imported] - Vibrio cholerae (group O1 strain N16961).
C:Species: Vibrio cholerae
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 15-Sep-2000
C:Accession: B82197
R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwin, M.L.; Dodson, R.J.;
Chardon, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, F.
l, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A:Reference number: A82035; MUID:20406833
A:Accession: B82197
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-399 <HEI>
A:Cross-references: GB:AE003852; NID:g9655952; PIDN:AAF94615.1; GSPDB:GN001
A:Experimental source: serogroup O1; strain N16961; biotype El Tor
C:Genetics:
A:Gene: VC1458
A:Map position: 1

Query Match          82.6%; Score 38; DB 2; Length 399;
Best Local Similarity 87.5%; Pred. No. 9.1;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGLCYVDG 8
    1 111111
DB 291 GRLCYVDG 298

RESULT 4
AA3864
zonula occludens toxin - Vibrio cholerae
C:Species: Vibrio cholerae
C:Date: 10-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 08-Oct-1999
C:Accession: A43864
R:Baudry, B.; Fasano, A.; Kettley, J.; Kaper, J.B.
Infect. Immun. 60, 428-434, 1992
A:Title: Cloning of a gene (zot) encoding a new toxin produced by Vibrio cholerae.
A:Reference number: A43864; MUID:92112300
A:Accession: A43864
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-399 <BAU>
A:Cross-references: GB:M83563; NID:g155314; PIDN:AAA27592.1; PID:g155315
A:Note: sequence extracted from NCBI backbone (NCBIN:77488, NCBIPI:77491)

Query Match          82.6%; Score 38; DB 2; Length 399;
Best Local Similarity 87.5%; Pred. No. 9.1;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGLCYVDG 8
    1 111111
DB 291 GRLCYVDG 298

RESULT 5
AA8825
Notch homolog Notch protein - mouse (fragment)
C:Species: Mus musculus (house mouse)
C:Date: 01-Dec-1993 #sequence_revision 18-Nov-1994 #text_change 13-Aug-1999

```

```

C:Accession: A48825
R:Reame, A.G.; Conlon, R.A.; Ziringibl, R.; Yamaguchi, T.P.; Rossant, J.
Dev. Biol. 154, 377-387, 1992
A:Title: Expression analysis of a Notch homologue in the mouse embryo.
A:Reference number: A48825; MUID:93050801
A:Accession: A48825
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-861 <REA>
A:Experimental source: embryo
A:Note: sequence extracted from NCBI backbone (NCBIPI:119144)
C:Superfamily: unassigned ankyrin repeat proteins; ankyrin repeat homology; EGF homol
F:26-57/Domain: EGF homology <EGF>
F:64-95/Domain: EGF homology <EGX1>
F:198-229/Domain: EGF homology <EGF2>
F:441-472/Domain: EGF homology <EGX2>

Query Match          80.4%; Score 37; DB 2; Length 861;
Best Local Similarity 75.0%; Pred. No. 29;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGLCYVDG 8
    1 111111
DB 327 GGLCYVDG 334

RESULT 6
AA6019
Notch-1 protein - mouse
N:Alternate names: notch protein
C:Species: Mus musculus (house mouse)
C:Date: 22-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 20-Sep-1999
C:Accession: A46019; S25144
R:del Amo, F.F.; Gendron-Maguire, M.; Swiatek, P.J.; Jenkins, N.A.; Copeland, N.G.; G
enomics 15, 259-264, 1993
A:Title: Cloning, analysis, and chromosomal localization of Notch-1, a mouse homolog
A:Reference number: A46019; MUID:93194170
A:Accession: A46019
A:Status: not compared with conceptual translation
A:Molecule type: nucleic acid
A:Residues: 1-2531 <DEU>
A:Cross-references: GB:Z11886; GB:S47228; NID:g288502; PIDN:CAA77941.1; PID:g288503
A:Note: sequence extracted from NCBI backbone (NCBIPI:127318)
R:Franco del Amo, F.; Smith, D.E.; Swiatek, P.J.; Gendron-Maguire, M.; Greenspan, R.J.
submitted to the EMBL Data Library, April 1992
A:Description: Expression pattern of Notch, a mouse homolog of Drosophila Notch, sug
A:Reference number: S25144
A:Accession: S25144
A:Molecule type: mRNA
A:Residues: 1551-2108, 'Q', 2110-2114, 'ALP', 2118-2170 <FRA>
A:Cross-references: EMBL:Z11886
C:Genetics:
A:Gene: notch-1
A:Map position: 2
A:Note: proximal region of chromosome 2
C:Superfamily: unassigned ankyrin repeat proteins; ankyrin repeat homology; EGF homol
F:106-138/Domain: EGF homology <EGF1>
F:144-175/Domain: EGF homology <EGF2>
F:222-254/Domain: EGF homology <EGF3>
F:261-292/Domain: EGF homology <EGF4>
F:339-370/Domain: EGF homology <EGF5>
F:416-449/Domain: EGF homology <EGF6>
F:456-487/Domain: EGF homology <EGF7>
F:494-525/Domain: EGF homology <EGF8>
F:532-563/Domain: EGF homology <EGF9>
F:607-638/Domain: EGF homology <EGF10>
F:682-713/Domain: EGF homology <EGF11>
F:757-788/Domain: EGF homology <EGF12>
F:795-826/Domain: EGF homology <EGF13>
F:873-904/Domain: EGF homology <EGF14>
F:911-942/Domain: EGF homology <EGF15>
F:949-980/Domain: EGF homology <EGF16>

```



A:Description: Involvement of a novel gene, xln, in cardiac looping.

A:Reference number: Z17948

A:Accession: T14267

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-1677 <MAN>

A:Cross-references: EMBL:AF051945; NID:g2970645; PID:g2970646; PIDN:AAC06023.1

A:Experimental source: cardiac muscle; stage early embryo

Query Match 76.1%; Score 35; DB 2; Length 1677;

Best Local Similarity 75.0%; Pred. No. 85;

Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGLCYODG 8

DB 886 GGLCYODG 893

RESULT 11

T26723

hypothetical protein Y39A1A.11 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 31-Jan-2000

C:Accession: T26723

R:Wall, M.

submitted to the EMBL Data Library, September 1998

A:Reference number: Z20257

A:Accession: T26723

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-236 <WILD>

A:Cross-references: EMBL:AL031633; PIDN:CAA21012.1; GSPDB:GN00021; CESP:Y39A1A.11

A:Experimental source: clone Y39A1A

C:Genetics:

A:Gene: CESP:Y39A1A.11

A:Map position: 3

A:Introns: 45/3; 63/1; 114/3; 143/3; 225/3

C:Superfamily: ribitol dehydrogenase; short-chain alcohol dehydrogenase homology

Query Match 76.1%; Score 35; DB 2; Length 296;

Best Local Similarity 75.0%; Pred. No. 26;

Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GGLCYODG 8

DB 16 GGLCYODG 23

RESULT 12

D83353

mannitol dehydrogenase PA2342 [Imported] - Pseudomonas aeruginosa (strain PA01)

C:Species: Pseudomonas aeruginosa

C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000

C:Accession: D83353

R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Br

adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lardig, K.; Lm,

.; Lory, S.; Olson, M.V.

Nature 406, 959-964, 2000

A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho

A:Reference number: AB2950; MUID:20437337

A:Accession: D83353

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-491 <STO>

A:Cross-references: GB:AE004660; GB:AE004091; NID:g9948372; PIDN:AAG05730.1; GSPDB:GN001

A:Experimental source: strain PA01

C:Genetics:

A:Gene: mtlD; PA2342

C:Superfamily: conserved hypothetical protein YEL070w

Query Match 76.1%; Score 35; DB 2; Length 491;

Best Local Similarity 62.5%; Pred. No. 41;

Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 GGLCYODG 8

DB 134 GGLCYODG 141

RESULT 13

T30620

hypothetical protein 18L - Molluscum contagiosum virus 1

N:Alternate names: MC018L

C:Species: Molluscum contagiosum virus 1

C:Date: 05-Nov-1999 #sequence\_revision 05-Nov-1999 #text\_change 11-May-2000

C:Accession: T30620

R:Senkevich, T.G.; Bugert, J.J.; Sisler, J.R.; Koonin, E.V.; Darai, G.; Moss, B.

Science 273, 813-816, 1996

A:Title: Genome sequence of a human tumorigenic poxvirus: Prediction of specific host

A:Reference number: Z20876; MUID:96325459

A:Accession: T30620

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-680 <SEN>

A:Cross-references: EMBL:U60315; PIDN:AAC5146.1.

C:Genetics:

A:Note: MC018L

Query Match 76.1%; Score 35; DB 2; Length 680;

Best Local Similarity 85.7%; Pred. No. 56;

Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GGLCYODG 8

DB 151 GGLCYODG 157

RESULT 14

S70397

zona pellucida glycoprotein A - dog

C:Species: Canis lupus familiaris (dog)

C:Date: 28-Oct-1996 #sequence\_revision 27-Feb-1997 #text\_change 20-Aug-1999

C:Accession: S70397

R:Harris, J.D.; Hilder, D.W.; Fontenot, G.K.; Hsu, K.T.; Yurewicz, E.C.; Sacco, A.G.

DNA Seq. 4, 361-393, 1994

A:Title: Cloning and characterization of zona pellucida genes and cDNAs from a variet

A:Reference number: S70396; MUID:95143578

A:Accession: S70397

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-715 <HAR>

A:Cross-references: EMBL:U05779; NID:g458274; PIDN:AAA74386.1; PID:g458275

C:Superfamily: sperm-binding glycoprotein ZP2; ZP domain homology

F:368-628/Domain: ZP domain homology <ZPH>

Query Match 76.1%; Score 35; DB 2; Length 715;

Best Local Similarity 75.0%; Pred. No. 59;

Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GGLCYODG 8

DB 366 GGLCYODG 373

RESULT 15

S70398

zona pellucida glycoprotein A - cat

C:Species: Felis silvestris catus (domestic cat)

C:Date: 28-Oct-1996 #sequence\_revision 27-Feb-1997 #text\_change 20-Aug-1999

C:Accession: S70398

R:Harris, J.D.; Hilder, D.W.; Fontenot, G.K.; Hsu, K.T.; Yurewicz, E.C.; Sacco, A.G.

DNA Seq. 4, 361-393, 1994  
A:Title: Cloning and characterization of zona pellucida genes and cDNAs from a variety of  
A:Reference number: S70396; MUID:95143578  
A:Accession: S70396  
A:Status: Preliminary  
A:Molecule type: mRNA  
A:Residues: 1-716 <HAR>  
A:Cross-references: EMBL:U05776; NID:9458268; PIDN:AA74388.1; PID:9458269  
C:Superfamily: sperm-binding glycoprotein ZP2; ZP domain homology  
F:370-630/Domain: ZP domain homology <ZPH>

Query Match 76.1%; Score 35; DB 2; Length 716;  
Best Local Similarity 75.0%; Pred. No. 59;  
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
OY 1 GGLCVQDG 8  
DB 368 GDLCTQDG 375

Search completed: June 13, 2001, 14:10:47  
Job time: 148 sec

**This Page Blank (uspto)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 13, 2001, 14:21:48 ; Search time 74.44 Seconds  
(without alignments)  
3.681 Million cell updates/sec

Title: PCT-US01-05825A-18  
Perfect score: 46  
Sequence: 1 GGCACVQDG 8

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SWISSPROT\_39:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	38	82.6	399	1 ZOT_VIBCH	P38442 vibrio chol
2	37	80.4	2531	1 NTC1_MOUSE	Q01705 mus musculu
3	36	78.3	788	1 TRS1_HCMVA	P09695 human cytom
4	36	78.3	846	1 TRS1_HCMVA	P09715 human cytom
5	35	76.1	715	1 ZP2_CANFA	P47983 canis famli
6	35	76.1	716	1 ZP2_FELCA	P47984 felis silve
7	35	76.1	860	1 BGL1_ASPAC	P46825 aspergillus
8	34	73.9	340	1 ASPG_FNAME	Q47898 flavobacter
9	34	73.9	355	1 END3_SCHPO	Q05907 schizosach
10	34	73.9	444	1 TRSH_HUMAN	P17752 homo sapien
11	34	73.9	444	1 TRSH_RABIT	P17290 oryctolagus
12	34	73.9	444	1 TRSH_RAT	P08810 rattus norv
13	34	73.9	447	1 TRSH_MOUSE	P17532 mus musculu
14	34	73.9	666	1 ZP2_RABIT	P46829 oryctolagus
15	34	73.9	745	1 ZP2_HUMAN	Q05996 homo sapien
16	34	73.9	745	1 ZP2_MACRA	O77726 macaca radi
17	34	73.9	1453	1 CA11_CHICK	P07457 gallus gall
18	34	73.9	1877	1 PCK3_MOUSE	Q04592 mus musculu
19	33.5	72.8	704	1 MEPR_MOUSE	Q05847 mus musculu
20	33	71.7	115	1 GUAN_HUMAN	Q07747 homo sapien
21	33	71.7	142	1 NCBI_YEAST	P40096 saccharomyc
22	33	71.7	266	1 LIZB_HAETN	Q05394 haemophilus
23	33	71.7	362	1 FIBP_ADE09	P36846 human adeno
24	33	71.7	379	1 WIF1_HUMAN	Q09545 homo sapien
25	33	71.7	379	1 WIF1_MOUSE	Q09541 mus musculu
26	33	71.7	473	1 FP2_MYTGA	Q25464 mytilus gal
27	33	71.7	571	1 APB3_MOUSE	O88888 mus musculu
28	33	71.7	575	1 APB3_HUMAN	O96018 homo sapien
29	33	71.7	587	1 URE1_CLOPE	P94669 clostridium
30	33	71.7	732	1 YH8K_YEAST	Q03254 saccharomyc
31	33	71.7	1120	1 DPOL_RCMVA	Q085428 rat cytolec
32	33	71.7	1964	1 NTC4_MOUSE	P31695 mus musculu
33	33	71.7	2318	1 NTC3_MOUSE	Q61982 mus musculu

34	33	71.7	2444	1 NTC1_HUMAN	P46531 homo sapien
35	32	69.6	122	1 YPFO_ECOLI	P76511 escherichia
36	32	69.6	282	1 Y765_HAETN	O57125 haemophilus
37	32	69.6	403	1 YCOA_SYNP7	P42460 synechococc
38	32	69.6	444	1 P44_PANTR	P27473 pan troglod
39	32	69.6	457	1 P44H_CAEL	P09025 caenorhabdi
40	32	69.6	722	1 Y022_TREPA	O83066 treponema p
41	32	69.6	724	1 RPRO_TNVD	P27209 tobacco nec
42	32	69.6	752	1 UCS3_PEA	Q43093 pisum sativ
43	32	69.6	825	1 BGLS_HANAN	P06835 hansecula
44	32	69.6	1153	1 BGL1_MOUSE	P52332 mus musculu
45	31	67.4	155	1 NEU1_FUGRU	O42493 fugu rubrip

## ALIGNMENTS

```

RESULT 1
ID ZOT_VIBCH STANDARD: PRT: 399 AA.
AC P38442: Q9L706: Q9R3V6:
DT 01-OCT-1994 (rel. 30, Created)
DT 01-OCT-2000 (rel. 40, Last sequence update)
DT 01-OCT-2000 (rel. 40, Last annotation update)
DE ZONA OCCLUDENS TOXIN (ZONULAR OCCLUDENS TOXIN).
GN ZOT OR VC1458.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX NCBI_TaxID=666;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CLASSICAL INABA 569B:
RX MEDLINE=92112300; PubMed=1730472;
RA Baudry B., Fasano A., Kelley J., Kaper J.B.:
RT "Cloning of a gene (zot) encoding a new toxin produced by Vibrio
RT cholerae."
RL Infect. Immun. 60:428-434(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=KNH002:
RX Shiu H.J., Park Y.C., Kim Y.C.:
RT "Cloning and nucleotide sequence analysis of the virulence gene
RT cassette from Vibrio cholerae KNH002 isolated in Korea."
RL Misamunrah Hoji 35:205-210(1999).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=O139-TOR OGAWA:
RX Zhi-Yong H., Wei-Jie Z., Xiang-Fu W.:
RT "Cloning and Expression of zot Gene from Vibrio cholerae."
RL Submitted (JAN-1999) to the EMBL/Genbank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=EL TOR 86015 / SEROTYPE O1:
RX Kan B., Liu Y.Q., Qi G.M., Gao S.Y.:
RT Submitted (JAN-2000) to the EMBL/Genbank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=EL TOR N16961 / SEROTYPE O1:
RX MEDLINE=20406833; PubMed=10952201;
RT Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwin M.L.,
RT Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RT Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RT Ermolaeva M.D., Yamathayan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA McDonald L.G., Uteerback T., Fleischmann R.D., Nierman W.C., White O.,
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA Fraser C.M.:
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT cholerae."
RL Nature 406:477-483(2000).
RN [6]
RP CHARACTERIZATION.
RX MEDLINE=91271365; PubMed=2052603;
RA Fasano A., Baudry B., Pumpilin D.W., Wasserman S.S., Tall B.D.,

```

RA Kelley J.M., Kaper J.B. ;  
RT "Vibrio cholerae produces a second enterotoxin, which affects  
RL intestinal tight junctions." ;  
RL Proc. Natl. Acad. Sci. U.S.A. 88:5242-5246(1991).  
CC -1- FUNCTION: INCREASES THE PERMEABILITY OF THE SMALL INTESTINE MUCOSA  
CC (ZONULA OCCUDENS).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb.ch/announce/>  
CC or send an email to [license@isb.ch](mailto:license@isb.ch)).  
CC -----  
DR EMBL; M83563; AAA27582.1; -  
DR EMBL; AF175708; AAD51358.1; -  
DR EMBL; AF123049; AAD26854.1; -  
DR EMBL; AF220606; AAF29547.1; -  
DR EMBL; AE004224; AAF94615.1; -  
DR PIR; A43864; A43864.  
DR TIGR; VC1458; -  
KM Enterotoxin; Toxin.  
FT VARIANT 45 M -> I (IN STRAIN 569B).  
FT VARIANT 100 V -> A (IN STRAINS 569B AND 86015).  
FT VARIANT 272 V -> A (IN STRAIN 569B).  
FT VARIANT 281 V -> A (IN STRAIN 569B).  
FT VARIANT 349 A -> S (IN STRAIN 86015).  
FT VARIANT 381 K -> R (IN STRAIN 86015).  
FT CONFLICT 386 IKTENDKKGLNSIF -> VKKEKEESLTKSL (IN REF.  
FT SEQUENCE 399 AA; 44903 MW; 3C7424B75816774 CRC64;  
SQ  
Query Match 82.6%; Score 38; DB 1; Length 399;  
Best Local Similarity 87.5%; Pred. No. 5.4;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 GGLGVQDG 8  
ID 111111  
Db 291 GRLGVQDG 298  
RESULT 2  
NTCL\_MOUSE STANDARD; PRT; 2531 AA.  
AC Q01705;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-FEB-1996 (Rel. 33, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 1 PRECURSOR (NOTCH PROTEIN).  
GN NOTCH1 OR NOTCH.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Embryo;  
RX MEDLINE=93194170; PubMed=8449489;  
RA Franco del Amo F., Gendron-Maguire M., Swiatek P.J., Jenkins N.A.,  
RA Copeland N.G., Gridley T. ;  
RT "Cloning, analysis, and chromosomal localization of Notch-1, a mouse  
RT homolog of Drosophila Notch." ;  
RL Genomics 15:259-264(1993).  
[2]  
RN [2]  
RP SEQUENCE OF 1551-2170 FROM N.A.  
RC TISSUE=Embryo;  
RX MEDLINE=9304835; PubMed=1425352;  
RA Franco del Amo F., Smith D.E., Swiatek P.J., Gendron-Maguire M.,  
RA Greenpan R.J., McMahon A.P., Gridley T. ;  
RT "Expression pattern of Notch, a mouse homolog of Drosophila Notch,"

[illegible]



```

FT DOMAIN 1145 1181 EGF-LIKE 30, CALCULUM-BINDING (POTENTIAL).
FT DOMAIN 1183 1219 EGF-LIKE 31, CALCULUM-BINDING (POTENTIAL).
FT DOMAIN 1221 1265 EGF-LIKE 32, CALCULUM-BINDING (POTENTIAL).
FT DOMAIN 1267 1305 EGF-LIKE 33.
FT DOMAIN 1307 1346 EGF-LIKE 34.
FT DOMAIN 1348 1384 EGF-LIKE 35.
FT DOMAIN 1387 1426 EGF-LIKE 36.
FT DOMAIN 1449 1462 CYS-RICH.
FT REPEAT 1445 1480 LIN/NORCH 1.
FT REPEAT 1481 1522 LIN/NORCH 2.
FT REPEAT 1523 1562 LIN/NORCH 3.
FT REPEAT 1562 1597 ANK 1.
FT REPEAT 1597 1979 ANK 2.
FT REPEAT 1949 1979 ANK 3.
FT REPEAT 1983 2012 ANK 4.
FT REPEAT 2016 2045 ANK 5.
FT REPEAT 2049 2078 ANK 5.
FT DISULFID 24 37 BY SIMILARITY.
FT DISULFID 31 46 BY SIMILARITY.
FT DISULFID 63 74 BY SIMILARITY.
FT DISULFID 68 87 BY SIMILARITY.
FT DISULFID 89 98 BY SIMILARITY.
FT DISULFID 106 117 BY SIMILARITY.
FT DISULFID 111 127 BY SIMILARITY.
FT DISULFID 129 138 BY SIMILARITY.
FT DISULFID 144 155 BY SIMILARITY.
FT DISULFID 149 164 BY SIMILARITY.
FT DISULFID 166 175 BY SIMILARITY.
FT DISULFID 182 195 BY SIMILARITY.
FT DISULFID 189 204 BY SIMILARITY.
FT DISULFID 206 215 BY SIMILARITY.
FT DISULFID 222 233 BY SIMILARITY.
FT DISULFID 227 243 BY SIMILARITY.
FT DISULFID 245 254 BY SIMILARITY.
FT DISULFID 261 272 BY SIMILARITY.
FT DISULFID 266 281 BY SIMILARITY.
FT DISULFID 283 292 BY SIMILARITY.
FT DISULFID 299 312 BY SIMILARITY.
FT DISULFID 306 321 BY SIMILARITY.
FT DISULFID 323 332 BY SIMILARITY.
FT DISULFID 339 350 BY SIMILARITY.
FT DISULFID 344 359 BY SIMILARITY.
FT DISULFID 361 370 BY SIMILARITY.
FT DISULFID 376 387 BY SIMILARITY.
FT DISULFID 381 398 BY SIMILARITY.
FT DISULFID 400 409 BY SIMILARITY.
FT DISULFID 416 429 BY SIMILARITY.
FT DISULFID 423 438 BY SIMILARITY.
FT DISULFID 440 449 BY SIMILARITY.
FT DISULFID 456 467 BY SIMILARITY.
FT DISULFID 461 476 BY SIMILARITY.
FT DISULFID 478 487 BY SIMILARITY.
FT DISULFID 494 505 BY SIMILARITY.
FT DISULFID 499 514 BY SIMILARITY.
FT DISULFID 516 525 BY SIMILARITY.
FT DISULFID 532 543 BY SIMILARITY.
FT DISULFID 537 552 BY SIMILARITY.
FT DISULFID 554 563 BY SIMILARITY.
FT DISULFID 560 580 BY SIMILARITY.
FT DISULFID 570 589 BY SIMILARITY.
FT DISULFID 575 589 BY SIMILARITY.
FT DISULFID 591 600 BY SIMILARITY.
FT DISULFID 607 618 BY SIMILARITY.
FT DISULFID 612 627 BY SIMILARITY.
FT DISULFID 629 638 BY SIMILARITY.
FT DISULFID 643 653 BY SIMILARITY.
FT DISULFID 650 664 BY SIMILARITY.
FT DISULFID 666 675 BY SIMILARITY.
FT DISULFID 682 693 BY SIMILARITY.
FT DISULFID 687 702 BY SIMILARITY.
FT DISULFID 704 713 BY SIMILARITY.
FT DISULFID 720 730 BY SIMILARITY.
FT DISULFID 725 739 BY SIMILARITY.
FT DISULFID 741 750 BY SIMILARITY.
FT DISULFID 757 768 BY SIMILARITY.

```

```

FT DISULFID 762 777 BY SIMILARITY.
FT DISULFID 779 788 BY SIMILARITY.
FT DISULFID 795 806 BY SIMILARITY.
FT DISULFID 800 815 BY SIMILARITY.
FT DISULFID 817 826 BY SIMILARITY.
FT DISULFID 833 844 BY SIMILARITY.
FT DISULFID 838 855 BY SIMILARITY.
FT DISULFID 857 866 BY SIMILARITY.
FT DISULFID 873 884 BY SIMILARITY.
FT DISULFID 878 893 BY SIMILARITY.
FT DISULFID 895 904 BY SIMILARITY.
FT DISULFID 911 922 BY SIMILARITY.
FT DISULFID 916 931 BY SIMILARITY.
FT DISULFID 933 942 BY SIMILARITY.
FT DISULFID 987 998 BY SIMILARITY.
FT DISULFID 992 1007 BY SIMILARITY.
FT DISULFID 1009 1018 BY SIMILARITY.
FT DISULFID 1025 1036 BY SIMILARITY.
FT DISULFID 1030 1045 BY SIMILARITY.
FT DISULFID 1047 1056 BY SIMILARITY.
FT DISULFID 1063 1074 BY SIMILARITY.
FT DISULFID 1068 1083 BY SIMILARITY.
FT DISULFID 1085 1094 BY SIMILARITY.
FT DISULFID 1101 1122 BY SIMILARITY.
FT DISULFID 1116 1131 BY SIMILARITY.
FT DISULFID 1133 1142 BY SIMILARITY.
FT DISULFID 1149 1160 BY SIMILARITY.

Query Match Score 37; DB 1; Length 2531;
Best Local Similarity 75.0%; Pred. No. 44;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```

```

Qy 1 GGLCVQDC 8
Db 1119 GGLCVDEG 1126

RESULT 3
TRSL_HCMVA STANDARD: PRT: 788 AA.
ID TRSL_HCMVA
AC P09695;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 01-JUN-1994 (Rel. 29, Last annotation update)
DE HYPOTHETICAL PROTEIN HHLF1.
GN TRSL.
OS Human cytomegalovirus (strain AD169).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Betaherpesvirinae; Cytomegalovirus.
OX NCBI_TaxID=10360;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87169717; PubMed=3031311;
RA Weston K., Barrell B.G.;
RT "Sequence of the short unique region, short repeats, and part of the
long repeats of human cytomegalovirus.";
RL J. Mol. Biol. 192:177-208(1986).
RN [2]
RP COMPLETE GENOME.
RX MEDLINE=90269039; PubMed=2161319;
RA Chee M.S., Bankier A.T., Beck S., Bohnt R., Brown C.M., Cerny R.,
Horsnell T., Hutchinson G.A. III, Kozakides T., Martignetti J.A.,
Predde E., Satchwell S.C., Tomlinson P., Weston K.M., Barrell B.G.;
RT "Analysis of the protein-coding content of the sequence of human
cytomegalovirus strain AD169.";
RL Curr. Top. Microbiol. Immunol. 154:125-169(1990).
RN [3]
RP -1- SIMILARITY: BELONGS TO THE US22 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```

entties requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: X17403; CAA35269.1; -

DR EMBL: X04650; CAB37121.1; -

DR PIR: C27349; OQBEB3.

DR PIR: S09951; S09951.

DR HSSP: P41249; IAKP.

KW Hypothetical protein.

FT CARBOHYD 76

FT CARBOHYD 118

FT CARBOHYD 223

SO SEQUENCE 788 AA; 83981 MW; 604BC69CA472BC7A CRC64;

Query Match

Best Local Similarity 85.7%; Score 36; DB 1; Length 788;

Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GLCVDG 8

Db 199 GLCVDG 205

RESULT 4

ID IRS1\_HCMVA STANDARD; PRT; 846 AA.

AC P09715;

DT 01-MAR-1989 (Rel. 10, Created)

DT 01-MAR-1989 (Rel. 10, Last sequence update)

DT 01-FEB-1991 (Rel. 17, Last annotation update)

DE HYPOTHETICAL PROTEIN HORE1.

OS IRS1.

GN Human cytomegalovirus (strain AD169).

OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;

OC Betaherpesvirinae; Cytomegalovirus.

OX NCBI\_TaxID=10360;

RA SEQUENCE FROM N.A.

RA MEDLINE=87169717; PubMed=3031311;

RA Weston K., Barrell B.G.;

RT "Sequence of the short unique region, short repeats, and part of the long repeats of human cytomegalovirus.";

RT J. Mol. Biol. 192:177-208 (1986).

RL [2]

RP COMPLETE GENOME.

RX MEDLINE=90269039; PubMed=2161319;

RA Chee M.S., Bankier A.T., Beck S., Brown C.M., Cerny R., Horsnell T., Hutchison C.A. III, Kuzarzides T., Martignetti J.A., Peddie E., Satchwell S.C., Tomlinson P., Weston K.M., Barrell B.G.;

RA "Analysis of the protein-coding content of the sequence of human cytomegalovirus strain AD169.";

RT Curr. Top. Microbiol. Immunol. 154:125-169 (1990).

RT -1- SIMILARITY: BELONGS TO THE US22 FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: X17403; CAA35311.1; -

DR EMBL: X04650; CAA28312.1; -

DR PIR: C26078; OQBEC3.

DR PIR: S09914; S09914.

DR HSSP: P41249; IAKP.

KW Hypothetical protein.

FT CARBOHYD 76

FT CARBOHYD 118

FT CARBOHYD 223

SO SEQUENCE 846 AA; 91048 MW; 622B6CFE211674BD CRC64;

Query Match

Best Local Similarity 85.7%; Score 36; DB 1; Length 846;

Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GLCVDG 8

Db 199 GLCVDG 205

RESULT 5

ID ZP2\_CANFA STANDARD; PRT; 715 AA.

AC P47983;

DT 01-FEB-1996 (Rel. 33, Created)

DT 01-FEB-1996 (Rel. 33, Last sequence update)

DT 01-FEB-1996 (Rel. 33, Last annotation update)

DE ZONA PELLUCCIDA SPERM-BINDING PROTEIN 2 PRECURSOR (ZONA PELLUCCIDA GLYCOPROTEIN ZP2) (ZONA PELLUCCIDA PROTEIN A).

DE ZP2 OR ZPA.

GN Canis familiaris (Dog).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.

OX NCBI\_TaxID=9615;

RA SEQUENCE FROM N.A.

RA TISSUE-Ovary;

RX MEDLINE=95143578; PubMed=7841460;

RA Harris J.D., Hibler D.W., Fontenot G.K., Hsu K.T., Yurewicz E.C., Sacco A.G.;

RT "Cloning and characterization of zona pellucida genes and cDNAs from a variety of mammalian species: the ZPA, ZPB and ZPC gene families.";

RT DNA Seq. 4:361-393 (1994).

RL [2]

RP SEQUENCE FROM N.A.

RC TISSUE-Ovary;

RA Okazaki Y., Isojima S., Sugimoto M.;

RA Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: U05779; AAA74386.1; -

DR EMBL: D45069; BAA08097.1; -

DR InterPro: IPR001507; -

DR Pfam: PF00100; zona\_pellucida; 1.

DR PRINTS: PR00023; ZPELUCCIDA.

DR PROSITE: PS00682; ZP\_DOMAIN; 1.

KW Glycoprotein; Signal; Sulfatation; Sperm; Receptor; Transmembrane; Extracellular matrix;

KM

FT SIGNAL 1

FT CHAIN 39

FT DOMAIN 39

FT TRANSMEM 685

FT DOMAIN 706

FT DOMAIN 367

FT CARBOHYD 87

FT CARBOHYD 193

FT CARBOHYD 220

FT CARBOHYD 266

FT CARBOHYD 321

FT CARBOHYD 321

FT CONFLICT 15

FT CONFLICT 292

FT CONFLICT 328

FT CONFLICT 599

SO SEQUENCE 715 AA; 79938 MW; 508B6DE804F4DC5C CRC64;

BY SIMILARITY.

ZP. ZONA PELLUCCIDA SPERM-BINDING PROTEIN 2.

EXTRACELLULAR (POTENTIAL).

POTENTIAL.

CYTOPLASMIC (POTENTIAL).

N-LINKED (GLCNAC. . .) (POTENTIAL).

CAR-LINKED (GLCNAC. . .) (POTENTIAL).

N-LINKED (GLCNAC. . .) (POTENTIAL).

N-LINKED (GLCNAC. . .) (POTENTIAL).

R -> W (IN REF. 2).

R -> A (IN REF. 2).

L -> P (IN REF. 2).

S -> A (IN REF. 2).

Query Match 76.1%; Score 35; DB 1; Length 715;  
 Best Local Similarity 75.0%; Pred. NO. 33;  
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 GGLCVODG 8  
 1 1 1 1 1  
 366 GDLCTODG 373

RESULT 6  
 ZP2\_FELCA STANDARD: PRT: 716 AA.  
 ID ZP2\_FELCA  
 AC P47984;  
 DT 01-FEB-1996 (Rel. 33, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE ZONA PELLUCIDA SPERM-BINDING PROTEIN 2 PRECURSOR (ZONA PELLUCIDA GLYCOPROTEIN ZP2) (ZONA PELLUCIDA PROTEIN A).  
 GN ZP2 OR ZPA.  
 OS Felis silvestris catus (Cat).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Carnivora; Fissipedia; Felidae; Felis.  
 OX NCBI\_Taxid=9685;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-Ovary;  
 RX MEDLINE=95143578; PubMed=7841460;  
 RA Harris J.D., Hibler D.W., Fontenot G.K., Hsu K.T., Yurewicz E.C., Sacco A.G.;  
 RT "Cloning and characterization of zona pellucida genes and cDNAs from a variety of mammalian species: the ZPA, ZPB and ZPC gene families";  
 RL DNA Seq. 4:361-393(1994).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-Ovary;  
 RX Okazaki Y., Isojima S., Sugimoto M.;  
 RA Submitted (JAN-1995) to the EMBL/Genbank/DBP databases.  
 RL -1- FUNCTION: ZP2 FORMS WITH ZP1 AND ZP3 THE ZONA PELLUCIDA, IN WHICH ZP2 AND ZP3 COMPLEX INTO COPOLYMERS CROSS-LINKED BY ZP1. ZP2 ACTS AS A SECONDARY SPERM RECEPTOR.  
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. EXTRACELLULAR MATRIX.  
 CC -1- PM: SULFATED GLYCOPROTEIN WITH O-LINKED OLIGOSACCHARIDES (BY SIMILARITY).  
 CC -1- SIMILARITY: CONTAINS 1 ZP DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL: U05776; AAA74388.1; -;  
 DR EMBL: D45067; BAA08095.1; -;  
 DR InterPro: IPR001507; -;  
 DR Pfam: PF00100; zona\_pellucida; 1.  
 DR PRINTS: PR00023; zpellucida.  
 DR PROSITE: PS00682; ZP\_DOMAIN: 1.  
 KW Glycoprotein; Signal; Sulfation; Sperm; Receptor; Transmembrane; Extracellular matrix.  
 FT SIGNAL 1 38  
 FT CHAIN 39 716 BY SIMILARITY.  
 FT DOMAIN 39 686 ZONA PELLUCIDA SPERM-BINDING PROTEIN 2.  
 FT TRANSMEM 687 707 EXTRACELLULAR (POTENTIAL).  
 FT DOMAIN 708 716 POTENTIAL.  
 FT DOMAIN 369 636 CYTOPLASMIC (POTENTIAL).  
 FT CARBOHYD 87 87 ZP.  
 FT CARBOHYD 96 96 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 222 222 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 268 268 N-LINKED (GLCNAC. . .) (POTENTIAL).

Query Match 76.1%; Score 35; DB 1; Length 716;  
 Best Local Similarity 75.0%; Pred. NO. 33;  
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 GGLCVODG 8  
 1 1 1 1 1  
 366 GDLCTODG 375

RESULT 7  
 BGL1\_ASPAC STANDARD: PRT: 860 AA.  
 ID BGL1\_ASPAC  
 AC P48825;  
 DT 01-FEB-1996 (Rel. 33, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE BETA-GLUCOSIDASE 1 PRECURSOR (EC 3.2.1.21) (GENTIOBIASE) (CELLOBIASE)  
 DE (BETA-D-GLUCOSIDE GLUCOHYDROLASE).  
 OS Aspergillus aculeatus.  
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes; Eurotiiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.  
 OX NCBI\_Taxid=5053;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=F-50;  
 RX MEDLINE=97082984; PubMed=8964516;  
 RA Kawaguchi T., Enoki T., Tsurumaki S., Sumitani J., Ueda M., Ooi T., Arai M.;  
 RT "Cloning and sequencing of the cDNA encoding beta-glucosidase 1 from Aspergillus aculeatus";  
 RL Gene 173:287-288(1996).  
 CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF TERMINAL, NON-REDUCING BETA-D-GLUCOSE RESIDUES WITH RELEASE OF BETA-D-GLUCOSE.  
 CC -1- PATHWAY: CELLULOSE DEGRADATION.  
 CC -1- SIMILARITY: BELONGS TO FAMILY 3 OF GLYCOSYL HYDROLASES.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL: D64088; BAA10968.1; -;  
 DR InterPro: IPR002772; -;  
 DR Pfam: PF00933; Glyco\_hydro\_3; 1.  
 DR Pfam: PF01913; Glyco\_hydro\_3\_C; 1.  
 DR PRINTS: PR00133; GLHYDRLASE3.  
 DR PROSITE: PS00775; GLYCOSYL\_HYDROL\_F3; 1.  
 KW Hydrolase; Glycosidase; Cellulose degradation; Glycoprotein; Signal.  
 FT SIGNAL 1 19  
 FT CHAIN 20 860 BETA-GLUCOSIDASE 1.  
 FT ACT\_SITE 280 280  
 FT CARBOHYD 61 61 BY SIMILARITY.  
 FT CARBOHYD 211 211 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 252 252 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 315 315 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 322 322 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 354 354 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 387 387 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 442 442 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 523 523 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 542 542 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 564 564 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 658 658 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 668 668 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 690 690 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 712 712 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 860 AA: 93052 MW: 4B484778B00FC694 CRC64;

Query Match 76.1%; Score 35; DB 1; Length 860;  
 Best Local Similarity 71.4%; Pred. No. 39;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGLCVOD 7  
 11:111  
 DB 86 GGMCIQD 92

RESULT 8  
 ASPG\_FLAME STANDARD: PRT: 340 AA.  
 ID ASPG\_FLAME  
 AC 047898;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE NM-(BETA-N-ACETYLGLUCOSAMINYL)-L-ASPARAGINASE PRECURSOR (EC 3.5.1.26)  
 DE (GLUCOSYLASPARAGINASE) (ASPARTYLGLUCOSAMINIDASE) (NA-(N-ACETYL-BETA-  
 DE GLUCOSAMINYL)-L-ASPARAGINE AMIDASE) (AGA).  
 OS Flavobacterium meningosepticum.  
 OC Bacteria; CF8 group; Flavobacteria; Flavobacteriaceae;  
 OC Chryseobacterium.  
 OX NCBI\_TaxID=238;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RC STRAIN-ELDER / ATCC 33958;  
 RX MEDLINE=95142655; Pubmed=7840643;  
 RA Tarentino A.L., Quinones G., Hauer C.R., Changchien L.M.,  
 RA Plummer T.H. Jr.;  
 RT "Molecular cloning and sequence analysis of Flavobacterium  
 RT meningosepticum glycosylasparaginase: a single gene encodes the alpha  
 RT Arch. Biochem. Biophys. 316:399-406(1995).  
 RL [2]  
 RP SEQUENCE FROM N.A. AND 197-211.  
 RX MEDLINE=94071939; Pubmed=8250923;  
 RA Tarentino A.L., Plummer T.H. Jr.;  
 RT "The first demonstration of a procaryotic glycosylasparaginase.";  
 RL Biochem. Biophys. Res. Commun. 197:179-186(1993).  
 RN [3]  
 RP X-RAY CRYSTALLOGRAPHY (2.32 ANGSTROMS).  
 RX MEDLINE=98200483; Pubmed=9541410;  
 RA Xian J., Tarentino A.L., Grimwood B.G., Plummer T.H. Jr., Cui T.,  
 RA Guan C., van Roey P.;  
 RT "Crystal structure of glycosylasparaginase from Flavobacterium  
 RT meningosepticum.";  
 RL Protein Sci. 7:774-781(1998).  
 CC -1- FUNCTION: CLEAVES THE GLCNAC-ASN BOND WHICH JOINS OLIGOSACCHARIDES  
 CC TO THE PEPTIDE OF ASPARAGINE-LINKED GLYCOPROTEINS. REQUIRES THAT  
 CC THE GLYCOSYLATED ASPARAGINE MOIETY IS NOT SUBSTITUTED ON  
 CC ITS AMINO (R1) AND CARBOXYL (R2) TERMINUS.  
 CC -1- CATALYTIC ACTIVITY: NA-(BETA-N-ACETYL-D-GLUCOSAMINYL)-L-ASPARAGINE  
 CC + H(2)O = N-ACETYL-BETA-GLUCOSAMINYLAMINE + L-ASPARATE.  
 CC -1- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA CHAINS.  
 CC -1- SIMILARITY: BELONGS TO THE ASPARAGINASE 2 FAMILY.  
 CC -1- SIMILARITY: BELONGS TO THE ASPARAGINASE 2 FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL: 008028; AAA68868.1; .

DR PDB: 1AVY; 29-APR-98.  
 DR MEROPS: T02.001; .  
 DR InterPro: IPR000246; .  
 DR Pfam: PF01112; Asparaginase\_2; 1.  
 KW Signal: Hydrolase; periplasmic; 3D-structure.  
 FT SIGNAL 1 45  
 FT CHAIN 46 196  
 FT CHAIN 197 340  
 FT DISULFID 113 126  
 FT DISULFID 213 277  
 FT ACT\_SITE 197 197  
 SQ SEQUENCE 340 AA: 37262 MW: 4C36E3061B4E3D7 CRC64;

Query Match 73.9%; Score 34; DB 1; Length 340;  
 Best Local Similarity 62.5%; Pred. No. 26;  
 Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 GGLCVOD 8  
 11:111  
 DB 312 GAYCICDG 319

RESULT 9  
 END3\_SCHPO STANDARD: PRT: 355 AA.  
 ID END3\_SCHPO  
 AC 009907;  
 DT 01-FEB-1996 (Rel. 33, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE ENDOUCLEASE III HOMOLOG (EC 4.2.99.18) (DNA-(APURINIC OR APYRIMIDINIC  
 DE SITE) LYASE).  
 GN NTH1 OR SPAC30D11.07.  
 OS Schizosaccharomyces pombe (fission yeast).  
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
 OC Schizosaccharomyces.  
 OX NCBI\_TaxID=4896;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=972;  
 RA Pearson D., Churcher C.M., Barrell B.G., Rajandream M.A., Walsh S.V.;  
 RL Submitted (NOV-1995) to the EMBL/Genbank/DBJ databases.  
 RN [2]  
 RP CHARACTERIZATION.  
 RX MEDLINE=96406995; Pubmed=8811082;  
 RA Roldan-Arjona T., Anselmino C., Lindahl T.;  
 RT "Molecular cloning and functional analysis of a Schizosaccharomyces  
 RT pombe homologue of Escherichia coli endonuclease III.";  
 RL Nucleic Acids Res. 24:3307-3312(1996).  
 CC -1- FUNCTION: HAS BOTH AN APURINIC AND/OR APYRIMIDINIC ENDOUCLEASE  
 CC ACTIVITY AND A DNA N-GLYCOSYLASE ACTIVITY. POSSESSES GLYCOSYLASE  
 CC ACTIVITY ON DIFFERENT TYPES OF DNA SUBSTRATES WITH PYRIMIDINES  
 CC DAMAGE. IT CAN RELEASE BOTH UREA AND THYMINE GLYCOL FROM DOUBLE-  
 CC STRANDED POLYMERS.  
 CC -1- COFACTOR: BINDS A 4FE-4S CLUSTER WHICH IS NOT IMPORTANT FOR THE  
 CC CATALYTIC ACTIVITY, BUT WHICH IS PROBABLY INVOLVED IN THE PROPER  
 CC POSITIONING OF THE ENZYME ALONG THE DNA STRAND (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE NTH/MUYR FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL: 267961; CA91893.1; .  
 DR InterPro: IPR001502; .  
 DR Pfam: PF00730; Endonuclease\_3; 1.  
 DR PROSITE: PS00764; ENDOUCLEASE\_III\_1; FALSE\_NEG.  
 DR PROSITE: PS01155; ENDOUCLEASE\_III\_2; 1.

KM DNA repair; Hydrolase; Glycosidase; Lyase; Iron-sulfur; 4Fe-4S.  
 FT METAL 210 210 IRON-SULFUR (4FE-4S) (BY SIMILARITY).  
 FT METAL 217 217 IRON-SULFUR (4FE-4S) (BY SIMILARITY).  
 FT METAL 220 220 IRON-SULFUR (4FE-4S) (BY SIMILARITY).  
 FT METAL 228 228 IRON-SULFUR (4FE-4S) (BY SIMILARITY).  
 SQ SEQUENCE 355 AA; 40226 MW; D2E8C93A0DFA0EB CRC64;

Query Match 73.9%; Score 34; DB 1; Length 355;  
 Best Local Similarity 71.4%; Pred. No. 27;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Oy 1 GGLCVD 7  
 Db 79 GGLCLED 85

RESULT 10  
 TRSH\_HUMAN STANDARD: PRT; 444 AA.  
 AC P17732; Q16736; (Rel. 15, Created)  
 DT 01-AUG-1990 (Rel. 15, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE TRYPTOPHAN 5-MONOOXYGENASE (EC 1.14.16.4) (TRYPTOPHAN 5-HYDROXYLASE).  
 GN TPH OR TRPH.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Carotina;  
 RX MEDLINE=90332431; PubMed=2377472;  
 RA Boulalard S., Darmon M.C., Ganem Y., Launay J.M., Mallet J.;  
 RT "Complete coding sequence of human tryptophan hydroxylase";  
 RL Nucleic Acids Res. 18:4257-4257(1990).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE=95077422; PubMed=7986090;  
 RA Tipper J.P., Clifton B.A., Ribeiro P., Kaufman S.;  
 RT "Cloning and expression of rabbit and human brain tryptophan  
 hydroxylase cDNA in Escherichia coli";  
 RL Arch. Biochem. Biophys. 315:445-453(1994).  
 CC -1- CATALYTIC ACTIVITY: L-TRYPTOPHAN + TETRAHYDROPTERIDINE + O(2) -  
 CC 5-HYDROXY-L-TRYPTOPHAN + DIHYDROPTERIDINE + H(2)O.  
 CC -1- COFACTOR: FERROUS ION.  
 CC -1- PATHWAY: THIS IS THE RATE-LIMITING ENZYME IN THE BIOSYNTHESIS OF  
 CC SEROTONIN IN THE CENTRAL NERVOUS SYSTEM AND CATALYZES THE FIRST  
 CC STEP OF THE SYNTHESIS OF MELANOTONIN IN THE PINAL GLAND.  
 CC -1- SUBUNIT: MULTIMER OF IDENTICAL SUBUNITS.  
 CC -1- SIMILARITY: BELONGS TO THE BIOPTERIN-DEPENDENT AROMATIC AMINO ACID  
 CC HYDROXYLASES FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL; X52836; CA37018.1; -;  
 DR EMBL; L29306; AAA67050.1; -;  
 DR PIR; S10489; S10489.  
 DR HSSP; P04177; ITOH.  
 DR MIM; 191060; -;  
 DR InterPro; IPR001273; -;  
 DR InterPro; IPR002912; -;  
 DR Pfam; PF01842; ACT; 1.  
 DR Pfam; PF00351; bioplerin\_H; 1.  
 DR PRINTS; PR00372; FWHYDRXLASE.

DR PROSITE; PS00367; BIOPTERIN\_HYDROXYL; 1.  
 KM Oxidoreductase; Monooxygenase; Serotonin biosynthesis; Iron;  
 FT MOD\_RES 58 58 PHOSPHORYLATION (BY CAPK) (POTENTIAL).  
 FT METAL 272 272 IRON (BY SIMILARITY).  
 FT METAL 277 277 IRON (BY SIMILARITY).  
 FT METAL 317 317 IRON (BY SIMILARITY).  
 FT CONFLICT 19 19 T -> S (IN REF. 1).  
 FT CONFLICT 68 68 T -> I (IN REF. 1).  
 FT CONFLICT 90 91 TP -> NL (IN REF. 1).  
 FT CONFLICT 97 97 M -> L (IN REF. 1).  
 FT CONFLICT 100 100 E -> D (IN REF. 1).  
 FT CONFLICT 104 104 S -> T (IN REF. 1).  
 FT CONFLICT 151 151 L -> S (IN REF. 2).  
 FT CONFLICT 154 154 N -> S (IN REF. 2).  
 FT CONFLICT 157 157 H -> Y (IN REF. 2).  
 FT CONFLICT 179 179 R -> Q (IN REF. 1).  
 FT CONFLICT 207 207 R -> Q (IN REF. 2).  
 FT CONFLICT 217 217 V -> I (IN REF. 2).  
 FT CONFLICT 344 344 A -> V (IN REF. 2).  
 FT CONFLICT 414 414 T -> A (IN REF. 2).  
 FT CONFLICT 419 419 S -> N (IN REF. 2).  
 FT CONFLICT 425 425 R -> Q (IN REF. 1).  
 FT CONFLICT 436 436 A -> G (IN REF. 2).  
 SQ SEQUENCE 444 AA; 51032 MW; 86C39884F72E120A CRC64;

Query Match 73.9%; Score 34; DB 1; Length 444;  
 Best Local Similarity 85.7%; Pred. No. 33;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 2 GGLCVDG 8  
 Db 319 GGLCVDG 325

RESULT 11  
 TRSH\_RABIT STANDARD: PRT; 444 AA.  
 AC P17290; Q29523;  
 DT 01-AUG-1990 (Rel. 15, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DE TRYPTOPHAN 5-MONOOXYGENASE (EC 1.14.16.4) (TRYPTOPHAN 5-HYDROXYLASE).  
 GN TPH.  
 OS Oryctolagus cuniculus (Rabbit).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.  
 OX NCBI\_TaxID=9986;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE=87289638; PubMed=3475690;  
 RA Grenet H.E., Ledley F.D., Reed L.L., Woo S.L.C.;  
 RT "Full-length cDNA for rabbit tryptophan hydroxylase: functional  
 RT domains and evolution of aromatic amino acid hydroxylases";  
 RL Proc. Natl. Acad. Sci. U.S.A. 84:5530-5534(1987).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE=95077422; PubMed=7986090;  
 RA Tipper J.P., Clifton B.A., Ribeiro P., Kaufman S.;  
 RT "Cloning and expression of rabbit and human brain tryptophan  
 hydroxylase cDNA in Escherichia coli";  
 RL Arch. Biochem. Biophys. 315:445-453(1994).  
 CC -1- CATALYTIC ACTIVITY: L-TRYPTOPHAN + TETRAHYDROPTERIDINE + O(2) -  
 CC 5-HYDROXY-L-TRYPTOPHAN + DIHYDROPTERIDINE + H(2)O.  
 CC -1- COFACTOR: FERROUS ION.  
 CC -1- PATHWAY: THIS IS THE RATE-LIMITING ENZYME IN THE BIOSYNTHESIS OF  
 CC SEROTONIN IN THE CENTRAL NERVOUS SYSTEM AND CATALYZES THE FIRST  
 CC STEP OF THE SYNTHESIS OF MELANOTONIN IN THE PINAL GLAND.  
 CC -1- SUBUNIT: MULTIMER OF IDENTICAL SUBUNITS.  
 CC -1- SIMILARITY: BELONGS TO THE BIOPTERIN-DEPENDENT AROMATIC AMINO ACID  
 CC HYDROXYLASES FAMILY.

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M17250; AAA31487.1; -
CC EMBL: L29305; AA67051.1; -
CC PIR: A32699; A32699.
CC HSSP: P04177; ITOH.
CC InterPro: IPR001273; -.
CC InterPro: IPR002912; -.
CC Pfam: PF01842; ACT. 1.
CC Pfam: PF00351; bioplerin.H; 1.
CC PRINTS: PR00372; FWHYDRYLASE.
CC PROSITE: PS00367; BIOPTERIN_HYDROXYL; 1.
CC Oxidoreductase; Monooxygenase; Serotonin biosynthesis; Iron;
CC Phosphorylation.
CC MOD_RES 58 58 PHOSPHORYLATION (BY CAPK) (POTENTIAL).
CC METAL 272 272 IRON (BY SIMILARITY).
CC METAL 277 277 IRON (BY SIMILARITY).
CC METAL 317 317 IRON (BY SIMILARITY).
CC CONFLICT 102 102 M -> L (IN REF. 1).
CC CONFLICT 151 151 L -> S (IN REF. 2).
CC CONFLICT 202 203 KY -> ND (IN REF. 1).
CC CONFLICT 207 207 R -> Q (IN REF. 2).
CC CONFLICT 390 390 T -> K (IN REF. 1).
CC SEQUENCE 444 AA; 5118 MW; BF182451B28ECD80 CRC64;

Query Match 73.9%; Score 34; DB 1; Length 444;
Best Local Similarity 85.7%; Pred. No. 33;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 GLCVDG 8
DB 319 GLCVDG 325

RESULT 12
TRSH_RAT STANDARD; PRT; 444 AA.
AC P09810;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE TRYPTOPHAN 5-MONOOXYGENASE (EC 1.14.16.4) (TRYPTOPHAN 5-HYDROXYLASE).
GN TPH.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Cranata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Wistar; Tissue=Pinel gland;
RA MEDLINE=88244702; PubMed=3379411;
RA Darnon M.C., Gilbert B., Levial V., Enret M., Maitre M., Mallet J.;
RT "Sequence of two mRNAs encoding active rat tryptophan hydroxylase.";
RL J. Neurochem. 51:312-316(1988).
RN [2]
RP SEQUENCE FROM N.A.
RC MEDLINE=91245924; PubMed=1645430;
RA Kim K.S., Wessel T.C., Stone D.M., Carver C.H., Joh T.H., Park D.H.;
RT "Molecular cloning and characterization of cDNA encoding tryptophan
RT hydroxylase from rat central serotonergic neurons.";
RL Brain Res. Mol. Brain Res. 9:277-283(1991).
RN [3]
RP SEQUENCE OF 167-261 FROM N.A.
RC MEDLINE=87005247; PubMed=2875901;
RA Darnon M.C., Grima B., Cash C.D., Maitre M., Mallet J.;

```

```

RT "Isolation of a rat pineal gland cDNA clone homologous to tyrosine
RT and phenylalanine hydroxylases.";
RL FEBS Lett. 206:43-46(1986).
CC -1- CATALYTIC ACTIVITY: L-TRYPTOPHAN + TETRAHYDROPTERIDINE + O(2) -
CC 5-HYDROXY-L-TRYPTOPHAN + DIHYDROPTERIDINE + H(2)O.
CC -1- COFACTOR: FERROUS ION.
CC -1- PATHWAY: THIS IS THE RATE-LIMITING ENZYME IN THE BIOSYNTHESIS OF
CC SEROTONIN IN THE CENTRAL NERVOUS SYSTEM AND CATALYZES THE FIRST
CC STEP OF THE SYNTHESIS OF MELATONIN IN THE PINEAL GLAND.
CC -1- SIMILARITY: BELONGS TO THE BIOPTERIN-DEPENDENT AROMATIC AMINO ACID
CC HYDROXYLASES FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M28000; AAA4282.1; -
CC EMBL: X53501; CAA37579.1; -.
CC PIR: J10034; WHRTW.
CC PIR: A24367; A24367.
CC PIR: A60034; A60034.
CC HSSP: P04177; ITOH.
CC InterPro: IPR001273; -.
CC InterPro: IPR002912; -.
CC Pfam: PF01842; ACT. 1.
CC Pfam: PF00351; bioplerin.H; 1.
CC PRINTS: PR00372; FWHYDRYLASE.
CC PROSITE: PS00367; BIOPTERIN_HYDROXYL; 1.
CC Oxidoreductase; Monooxygenase; Serotonin biosynthesis; Iron;
CC Phosphorylation.
CC MOD_RES 58 58 PHOSPHORYLATION (BY CAPK) (POTENTIAL).
CC METAL 272 272 IRON (BY SIMILARITY).
CC METAL 277 277 IRON (BY SIMILARITY).
CC METAL 317 317 IRON (BY SIMILARITY).
CC SEQUENCE 444 AA; 51068 MW; C3CF5245727CC825 CRC64;

Query Match 73.9%; Score 34; DB 1; Length 444;
Best Local Similarity 85.7%; Pred. No. 33;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 GLCVDG 8
DB 319 GLCVDG 325

RESULT 13
TRSH_MOUSE STANDARD; PRT; 447 AA.
AC P17532;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE TRYPTOPHAN 5-MONOOXYGENASE (EC 1.14.16.4) (TRYPTOPHAN 5-HYDROXYLASE).
GN TPH.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Cranata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=90243261; PubMed=2110547;
RA Stoll J., Kozak C.A., Goldman D.;
RT "Characterization and chromosomal mapping of a cDNA encoding
RT tryptophan hydroxylase from a mouse mastocytoma cell line.";
RL Genomics 7:88-96(1990).
CC -1- CATALYTIC ACTIVITY: L-TRYPTOPHAN + TETRAHYDROPTERIDINE + O(2) -
CC 5-HYDROXY-L-TRYPTOPHAN + DIHYDROPTERIDINE + H(2)O.

```





RA Deslattes Mays A., Cao Y., Xu R.X., Kang H.-L., Mitchell S.,  
RA Elchler E.E., Harris P.C., Venter J.C., Adams W.D.,  
RT "Genome duplications and other features in 12 Mb of DNA sequence from  
human chromosome 16p and 16q.";  
RL Genomics 60:295-308(1999).  
CC -1- FUNCTION: ZP2 FORMS WITH ZP1 AND ZP3 THE ZONA PELLUCIDA, IN  
CC WHICH ZP2 AND ZP3 COMPLEX INTO COPOLYMERS CROSS-LINKED BY ZP1.  
CC ZP2 ACTS AS A SECONDARY SPERM RECEPTOR.  
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. EXTRACELLULAR  
CC MATRIX.  
CC -1- TISSUE SPECIFICITY: OOCYTES.  
CC -1- DEVELOPMENTAL STAGE: EXPRESSED DURING THE 2-WEEK GROWTH PHASE OF  
CC OOGENESIS, PRIOR TO OVULATION.  
CC -1- PTM: IS PROTEOLYTICALLY CLEAVED AFTER FERTILIZATION, AND THIS  
CC MODIFICATION, ALONG WITH PRESUMED CHANGES IN ZP3 MAY PLAY AN  
CC IMPORTANT ROLE IN THE POSTFERTILIZATION BLOCK TO POLYSPERMY.  
CC -1- PTM: SULFATED GLYCOPROTEIN WITH O-LINKED OLIGOSACCHARIDES  
CC (BY SIMILARITY).  
CC -1- SIMILARITY: CONTAINS 1 ZP DOMAIN.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: M90366; AAA61335.1; -  
DR EMBL: AF001550; AAB67599.1; -  
DR PIR: A48833; A48833.  
DR MIM: 182888; -  
DR InterPro: IPR001507; -  
DR Pfam: PF00100; zona\_pellucida: 1.  
DR PRINTS: PR00023; ZPELUCIDA.  
DR PROSITE: PS00682; ZP\_DOMAIN: 1.  
KW Glycoprotein; Signal; Sulfatation; Sperm; Receptor; Transmembrane;  
KW Extracellular matrix.  
SQ  
FT SIGNAL 1 38  
FT CHAIN 39 745  
FT DOMAIN 39 716 ZONA PELLUCIDA SPERM-BINDING PROTEIN 2.  
FT DOMAIN 39 736 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 717 736 POTENTIAL.  
FT DOMAIN 737 745 CYTOPLASMIC (POTENTIAL).  
FT DOMAIN 370 637 ZP.  
FT CARBOHYD 87 87 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 105 105 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 122 122 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 223 223 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 269 269 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 400 400 N-LINKED (GLCNAC. . .) (POTENTIAL).  
SQ SEQUENCE 745 AA; 82356 MW; 2614DA79738F2CDD CRC64;

Query Match 73.9%; Score 34; DB 1; Length 745;  
Best Local Similarity 75.0%; Pred. No. 53;  
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 GGLCVQDG 8  
| | | | |  
DB 369 GELCTQDG 376

Search completed: June 13, 2001, 14:21:48  
Job time: 808 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

## OM protein - protein search, using sw model

Run on: June 13, 2001, 14:20:36 ; Search time 225.85 Seconds

(Without alignments)  
4.152 Million cell updates/sec

Title: PCT-US01-05825A-18  
Perfect score: 46  
Sequence: 1 GGLCVQDG 8

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 374700 seqs, 117207915 residues

Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

## Database :

SPREMBL\_15:\*

- 1: sp\_archaea:\*
- 2: sp\_bacteria:\*
- 3: sp\_fungi:\*
- 4: sp\_human:\*
- 5: sp\_invertebrate:\*
- 6: sp\_mammal:\*
- 7: sp\_mhc:\*
- 8: sp\_organelle:\*
- 9: sp\_phage:\*
- 10: sp\_plant:\*
- 11: sp\_protist:\*
- 12: sp\_unclassified:\*
- 13: sp\_vertebrate:\*
- 14: sp\_virus:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	41	89.1	196	2	084353 chlamydia t
2	41	89.1	196	2	09PK45 chlamydia m
3	38	82.6	323	2	09L8F5 vibrio mimi
4	38	82.6	399	2	09R3V6 vibrio chol
5	38	82.6	399	2	09L706 vibrio chol
6	37	80.4	793	3	000903 geannomy
7	37	80.4	861	11	090W58 mus sp. mot
8	36	78.3	633	10	09LZU6 arabidopsi
9	36	78.3	1677	11	070373 mus muscu
10	35	76.1	217	4	014964 homo sapien
11	35	76.1	549	5	09XX28 caenorhadi
12	35	76.1	549	11	09WU64 mus muscu
13	35	76.1	561	5	09V4M1 drosophila
14	35	76.1	640	14	011308 molluscum c
15	35	76.1	680	14	098187 molluscum c
16	35	76.1	955	4	099466 homo sapien
17	35	76.1	1999	4	099940 homo sapien
18	35	76.1	2003	4	000306 homo sapien
19	34	73.9	103	4	099227 homo sapien

20	34	73.9	119	6	077753 canis famli
21	34	73.9	166	2	069700 mycobacteri
22	34	73.9	233	1	09V241 pyrococcus
23	34	73.9	283	5	09NB39 plasmodium
24	34	73.9	354	11	09JJS1 mus muscu
25	34	73.9	373	11	090YV1 ratius norv
26	34	73.9	380	4	013086 homo sapien
27	34	73.9	450	5	096947 geodia cydo
28	34	73.9	488	5	09NB47 plasmodium
29	34	73.9	492	5	09NAW4 plasmodium
30	34	73.9	1031	10	065500 arabidopsi
31	34	73.9	1487	4	014047 homo sapien
32	34	73.9	1491	13	091718 xenopus lae
33	34	73.9	1548	11	062040 mus muscu
34	34	73.9	1722	5	019350 caenorhadi
35	34	73.9	2108	2	053901 mycobacteri
36	33	71.7	138	13	09PSF8 gallus gall
37	33	71.7	161	2	048805 legionella
38	33	71.7	172	13	002529 oryzias lat
39	33	71.7	243	5	018926 caenorhadi
40	33	71.7	256	5	022208 caenorhadi
41	33	71.7	273	4	09UHF1 homo sapien
42	33	71.7	274	1	09V1H6 pyrococcus
43	33	71.7	336	14	068405 human cytom
44	33	71.7	365	14	064823 human adeno
45	33	71.7	365	14	064822 human adeno

## ALIGNMENTS

RESULT	ID	PRELIMINARY:	PRT:	196 AA.
084353	084353	PRELIMINARY:	PRT:	196 AA.
AC	084353:			
DT	01-NOV-1998 (TREMBLrel. 08, Created)			
DT	01-NOV-1998 (TREMBLrel. 08, Last sequence update)			
DT	01-MAY-2000 (TREMBLrel. 13, Last annotation update)			
DE	MAF-TYPE PROTEIN.			
GN	MAF.			
OS	Chlamydia trachomatis.			
OC	Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.			
OX	NCBI_TaxID=813;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=D/UW-3/CX;			
RX	MEDLINE=99000809; PubMed=9784136;			
RA	Stephens R.S., Kaiman S., Lammell C.J., Fan J., Marathe R., Aravind L., Mitchell W.P., Olinger L., Tetusov R.L., Zhao Q., Koonin E.V.,			
RA	Davis R.W.,			
RT	"Genome sequence of an obligate intracellular pathogen of humans:			
RT	Chlamydia trachomatis."			
RL	Science 282:754-759(1998).			
DR	EMBL: AE001308; AAC67944.1; "			
DR	SEQUENCE 196 AA: 21978 MW: EDB6CFD52F93073 CRC64;			
QY	1 GGLCVQDG 8			
DB	150 GGLCVQDG 157			
Query Match	89.1%; Score 41; DB 2; Length 196;			
Best Local Similarity	87.5%; Pred. No. 1.4;			
Matches	7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;			
RESULT	2			
ID	09PK45	PRELIMINARY:	PRT:	196 AA.
AC	09PK45:			
DT	01-OCT-2000 (TREMBLrel. 15, Created)			
DT	01-OCT-2000 (TREMBLrel. 15, Last sequence update)			
DT	01-OCT-2000 (TREMBLrel. 15, Last annotation update)			

DE MAF PROTEIN.  
 GN TC0628.  
 OS Chlamydia muridarum.  
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
 OX NCBI\_TaxID=83560;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-MOPN / NIGG.  
 RX MEDLINE=20150255; PubMed=10684935;  
 RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,  
 White O., Hickey E.K., Peterson J., Uterback T., Berry K., Bass S.,  
 Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,  
 Gwin M., Nelson W., Deboy R., Kolonay J., McClarty G., Salzberg S.L.,  
 Eisen J., Fraser C.M.;  
 RT "Genome sequences of Chlamydia trachomatis MOPN and Chlamydia  
 pneumoniae AR39."  
 RN Nucleic Acids Res. 28:1397-1406(2000).  
 RP [2]  
 RC SEQUENCE FROM N.A.  
 RA Read T.D., Brunham R., Shen C., Gill S.R., Heidelberg J.F., White O.,  
 Hickey E.K., Peterson J., Unayam L.A., Uterback T., Berry K.,  
 Bass S., Linher K., Weidman J., Khouri H., Craven B., Dodson R.,  
 Dodson R., Gwin M., Nelson W., Deboy R., Kolonay J., McClarty G.,  
 Salzberg S.L., Eisen J., Fraser C.M.;  
 RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RL EMBL; AE002331; AAF39457.1; -  
 DR TIGR; TC0628; -  
 SQ SEQUENCE 196 AA; 21730 MW; 9B86B059E52D78A5 CRC64;

Query Match 89.1%; Score 41; DB 2; Length 196;  
 Best Local Similarity 87.5%; Pred. No. 1.4;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GGLCYVDG 8  
 |||||  
 DB 150 GGLCYVDG 157

RESULT 3  
 O9L8F5 PRELIMINARY; PRT; 323 AA.  
 AC O9L8F5;  
 DT 01-OCT-2000 (TREMblrel. 15, Created)  
 DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)  
 DT 01-OCT-2000 (TREMblrel. 15, Last annotation update)  
 DE ZOT (FRAGMENT).  
 GN ZOT.  
 OS Vibrio mimicus.  
 OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.  
 OX NCBI\_TaxID=674;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-PF5;  
 RX MEDLINE=20143766; PubMed=10678967;  
 RA Boyd E.F., Moyer K.E., Shi L., Waldor M.K.;  
 RT "Infectious CTXphi and the vibrio pathogenicity island prophage in  
 RT vibrio mimicus: evidence for recent horizontal transfer between V.  
 RT mimicus and V. cholerae."  
 RL Infect. Immun. 68:1507-1513(2000).  
 DR EMBL; AF207857; AAF40142.1; -  
 FT NON\_TER 1  
 FT NON\_TER 323  
 DE SEQUENCE 323 AA; 36306 MW; 01C12DAE9B873C3B CRC64;

Query Match 82.6%; Score 38; DB 2; Length 323;  
 Best Local Similarity 87.5%; Pred. No. 8.9;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GGLCYVDG 8  
 |||||  
 DB 240 GGLCYVDG 247

RESULT 4  
 O9R3V6 PRELIMINARY; PRT; 399 AA.  
 AC O9R3V6;  
 DT 01-MAY-2000 (TREMblrel. 13, Created)  
 DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)  
 DT 01-OCT-2000 (TREMblrel. 15, Last annotation update)  
 DE ZONULAR OCCUDENS TOXIN (ZONA OCCUDENS TOXIN).  
 GN ZOT OR VC1458.  
 OS Vibrio cholerae.  
 OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.  
 OX NCBI\_TaxID=666;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-KNH002;  
 RA Shin H.J., Park Y.C., Kim Y.C.;  
 RT "Cloning and nucleotide sequence analysis of the virulence gene  
 RT cassette from Vibrio cholerae KNH002 isolated in Korea."  
 RL Misaiimurag Holjl 35:205-210(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-O139-TOR OGAMA;  
 RA Zhi-Yong H., Wei-Jie Z., Xiang-Fu W.;  
 RT "Cloning and Expression of zot Gene from Vibrio cholerae."  
 RT Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-EL TOR N16961 / SEROTYPE O1;  
 RX MEDLINE=20406833; PubMed=10952301;  
 RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwin M.L.,  
 Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Unayam L.A.,  
 Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,  
 Ermolaeva M.D., Yamthavayan J., Bass S., Qin H., Dragoi I., Sellers P.,  
 McDonald L., Uterback T., Fleischmann R.D., Nierman W.C., White O.,  
 Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,  
 Fraser C.M.;  
 RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio  
 RT cholerae."  
 RL Nature 406:477-483(2000).  
 DR EMBL; AF175708; AAD51358.1; -  
 DR EMBL; AF123049; AAD26854.1; -  
 DR EMBL; AE004224; AAF94615.1; -  
 DR TIGR; VC1458; -  
 SQ SEQUENCE 399 AA; 44903 MW; 3C742AB758176774 CRC64;

Query Match 82.6%; Score 38; DB 2; Length 399;  
 Best Local Similarity 87.5%; Pred. No. 11;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GGLCYVDG 8  
 |||||  
 DB 291 GGLCYVDG 298

RESULT 5  
 O9L7O6 PRELIMINARY; PRT; 399 AA.  
 AC O9L7O6;  
 DT 01-OCT-2000 (TREMblrel. 15, Created)  
 DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)  
 DT 01-OCT-2000 (TREMblrel. 15, Last annotation update)  
 DE ZOT.  
 GN ZOT.  
 OS Vibrio cholerae.  
 OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.  
 OX NCBI\_TaxID=666;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-86015;  
 RA Kan B., Liu Y.Q., Qi G.M., Gao S.Y.;

RT "Vibrio cholerae nct-CTXphI whole genome, include rstr(RSTR),  
 RT rstr(Rstr), rstr(Rstr), cep(Cep), orfU(Orfu), ace(Ace) and zot(Zot)  
 RT genes."  
 RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF220606; AAF29547.1; -  
 SQ SEQUENCE 399 AA; 44990 MW; CFE63DBCCE23EE1 CRC64;

Query Match 82.6%; Score 38; DB 2; Length 399;  
 Best Local Similarity 87.5%; Pred. No. 11;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GGLCYODG 8  
 DB 291 GGLCYODG 298

RESULT 6  
 ID 000903 PRELIMINARY; PRT; 793 AA.  
 AC 000903;  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)  
 DE AVENACINASE.  
 OS Gaemannomyces graminis.  
 OC Eukaryota; Fungi; Ascomycota; Ascomycota Incertae sedis;  
 OC Magnaporthaceae; Gaemannomyces.  
 OX NCBI\_TaxID=29850;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=96112750; Pubmed=8664505;  
 RA Osbourn A., Bowyer P., Lunness P., Clarke B., Daniels M.;  
 RT "Fungal pathogens of oat roots and tomato leaves employ closely  
 RT related enzymes to detoxify different host plant saponins."  
 RL Mol. Plant Microbe Interact. 8:971-978(1995).  
 DR EMBL: U35463; AAB09777.1; -  
 DR MENDEL: 20854; Gaegr:3059;20854.  
 DR INTERPRO: IPR001764; -  
 DR INTERPRO: IPR002772; -  
 DR PFAM: PF00933; Glyco\_hydro\_3; 1.  
 DR PFAM: PF01915; Glyco\_hydro\_3\_C; 1.  
 DR PRINTS: PR00133; GLHYDRASE3.  
 DR PROSITE: PS00775; GLYCOSYL\_HYDROL\_F3; 1.  
 SQ SEQUENCE 793 AA; 85243 MW; E656A5DED76F8169 CRC64;

Query Match 80.4%; Score 37; DB 3; Length 793;  
 Best Local Similarity 85.7%; Pred. No. 34;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 GGLCYODG 8  
 DB 89 GGLCYODG 95

RESULT 7  
 ID 090W58 PRELIMINARY; PRT; 861 AA.  
 AC 090W58;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)  
 DE NOTCH-NOTCH PRODUCT HOMOLOG.  
 OS Mus sp.  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10095;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=93050801; Pubmed=1426644;  
 RA Reaume A.G., Conlon R.A., Ziringibl R., Yamaguchi T.P., Rossant J.;  
 RT "Expression analysis of a Notch homologue in the mouse embryo.";

RL Dev. Biol. 154:377-387(1992).  
 DR HSSP: P00740; IIXA.  
 DR INTERPRO: IPR000152; -  
 DR INTERPRO: IPR000561; -  
 DR INTERPRO: IPR000742; -  
 DR INTERPRO: IPR000800; -  
 DR INTERPRO: IPR001438; -  
 DR INTERPRO: IPR001881; -  
 DR INTERPRO: IPR002049; -  
 DR PFAM: PF00008; EGF\_14.  
 DR PFAM: PF00066; notch; 3.  
 DR PRINTS: PR00010; EGFLOOD.  
 DR PRINTS: PR00011; EGFLAMIN.  
 DR PROSITE: PS00020; ASX\_HYDROXYL; 7.  
 DR PROSITE: PS00022; EGF\_1; UNKNOWN\_13.  
 DR PROSITE: PS01186; EGF\_2; 9.  
 DR PROSITE: PS01187; EGF\_CA; 7.  
 SQ SEQUENCE 861 AA; 92706 MW; 84BAEB915C2671EB CRC64;

Query Match 80.4%; Score 37; DB 11; Length 861;  
 Best Local Similarity 75.0%; Pred. No. 36;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 GGLCYODG 8  
 DB 327 GGLCYODG 334

RESULT 8  
 ID 09LZ06 PRELIMINARY; PRT; 633 AA.  
 AC 09LZ06;  
 DT 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)  
 DE HYPOTHETICAL 71.7 KDA PROTEIN.  
 GN F16L2\_50.  
 GN Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;  
 OC Brassicales; Brassicaceae; Arabidopsie.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Jordan N., Bangert S., Wiedelmann R., Voss H., Unsel M., Mewes R.W.,  
 RA Rudd S., Lemcke K., Mayer K.F.X., Quetier F., Salanoubat M.;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA EU Arabidopsis sequencing project;  
 RL Submitted (Apr-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AL162459; CAB82808.1; -  
 KW Hypothetical protein.  
 SQ SEQUENCE 633 AA; 71695 MW; 2758B9C38381DF14 CRC64;

Query Match 78.3%; Score 36; DB 10; Length 633;  
 Best Local Similarity 75.0%; Pred. No. 42;  
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 GGLCYODG 8  
 DB 624 GGLCYODG 631

RESULT 9  
 ID 070373 PRELIMINARY; PRT; 1677 AA.  
 AC 070373;  
 DT 01-AUG-1998 (TREMBLrel. 07, Created)  
 DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)

DE XIN.  
 GN XIN.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 CX NCB1\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=CARDIAC MUSCLE;  
 RA Wang D.-Z., Hu X., Lin J.L.-C., Kitten G.T., Solursh M., Lin J.J.-C.;  
 RL Front. Biosci. 1:0-0(1996).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=CARDIAC MUSCLE;  
 RA Wang D.-Z., Lin J.J.-C.;  
 RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF051945; AAC06023.1; -.  
 DR MGI: MGI:1333878; Xln.  
 SQ SEQUENCE 1677 AA; 182084 MW; A201CFC9A710C7FF CRC64;

Query Match 78.3%; Score 36; DB 11; Length 1677;  
 Best Local Similarity 75.0%; Pred. NO. 1,1e+02;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 GGLCYODG 8  
 DB 886 GGLCYODG 893

RESULT 10

014964 PRELIMINARY; PRT; 217 AA.

AC 014964;  
 DT 01-NOV-1996 (TREMBlrel. 01, Created)  
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)  
 DE RAB-RELATED GTP-BINDING PROTEIN (FRAGMENT).  
 OS Homo sapiens (Human).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 CX NCB1\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=97237046; PubMed=9119394;  
 RA Stankovic T., Byrd P.J., Cooper P.R., McConville C.M., Munroe D.J.,  
 RA Riley J.H., Watts G.D.J., Ambrose H., McGuire G., Smith A.D.,  
 RA Sutcliffe A., Mills T., Taylor A.M.R.;  
 RT Construction of a transcription map around the gene for ataxia  
 telangiectasia: identification of at least four novel genes.;  
 RL Genomics 40:267-276(1997).  
 DR EMBL: X99962; CAA68227.1; -.  
 DR HSSP: P05713; 3RAB.  
 DR INTERPRO: IPR001806; -.  
 DR PFAM: PF00071; Ras; 1.  
 DR PRINTS: PRO0449; RASTRNSFRNG.  
 FT NON\_TER 1  
 FT NON\_TER 217  
 SQ SEQUENCE 217 AA; 24869 MW; AA9A29731F42B8F5 CRC64;

Query Match 76.1%; Score 35; DB 4; Length 217;  
 Best Local Similarity 62.5%; Pred. No. 23;  
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 GGLCYODG 8  
 DB 182 GGLCYODG 189

RESULT 11  
 09XX28 PRELIMINARY; PRT; 296 AA.

AC 09XX28;  
 DT 01-NOV-1999 (TREMBlrel. 12, Created)  
 DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)  
 DE Y39A1A.11 PROTEIN.  
 GN Y39A1A.11  
 OS Caenorhabditis elegans.  
 CC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 CC Rhabditidae; Peloderinae; Caenorhabditis.  
 CX NCB1\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=94150718; PubMed=7906398;  
 RA Wall M.;  
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
 elegans"; (SEP-1998) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=94150718; PubMed=7906398;  
 RA Wilson R., Alnsough R., Anderson K., Baynes C., Berks M.,  
 RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,  
 RA Craxton M., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,  
 RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,  
 RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,  
 RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,  
 RA Smaildon N., Smith A., Sonhammer E., Staden R., Sulston J.,  
 RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,  
 RA Watson A., Weinstock L., Wilkinson-Sproat J., Woldman P.;  
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
 elegans";  
 RL Nature 368:32-38(1994).  
 DR EMBL: AL031633; CAA21012.1; -.  
 DR HSSP: P19992; 1HDC.  
 DR INTERPRO: IPR002198; -.  
 DR INTERPRO: IPR002347; -.  
 DR INTERPRO: IPR002424; -.  
 DR PFAM: PF00106; adh\_short; 1.  
 DR PFAM: PF00678; adh\_short\_C2; 1.  
 DR PRINTS: PRO0080; SDRFAMILY.  
 DR PRINTS: PRO0081; GDRDH.  
 DR PRINTS: PRO1167; INSADHEMITY.  
 DR PROSITE: PS00061; ADH\_SHORT; UNKNOWNL1.  
 SQ SEQUENCE 296 AA; 31244 MW; B1518A53D4ABE025 CRC64;

Query Match 76.1%; Score 35; DB 5; Length 296;  
 Best Local Similarity 75.0%; Pred. No. 32;  
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 GGLCYODG 8  
 DB 16 GGLCYODG 23

RESULT 12

09W064 PRELIMINARY; PRT; 549 AA.

AC 09W064;  
 DT 01-NOV-1999 (TREMBlrel. 12, Created)  
 DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)  
 DE GLYCEROL KINASE-LIKE PROTEIN 1.  
 GN GK-RS1 OR GKRS1.  
 OS Mus musculus (Mouse).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 CX NCB1\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=129SV; TISSUE=TESTES;  
 RA Pan Y., Decker W.K., Hug A.H.H.M., Cralgen W.J.;

"Retrotransposition of glycerol kinase-related genes from the X chromosome to autosomes: Functional and evolutionary aspects.";  
 RT Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.  
 RA EMBL; AF117733; AAD24550.1; -  
 DR HSSP; P08859; 1GLC.  
 DR MGI; 891990; GK-rs1.  
 DR INTERPRO: IPR000577; -  
 DR INTERPRO: IPR002086; -  
 DR PFM; PF00370; FGGY; 1.  
 DR PROSITE: PS00070; ALDEHYDE\_DEHYDR\_CYS; UNKNOWN\_1.  
 DR PROSITE: PS00445; FGGY\_KINASES\_2; 1.  
 DR PROSITE: PS00933; FGGY\_KINASES\_1; UNKNOWN\_1.  
 KW Kinase.  
 SO SEQUENCE 549 AA; 59871 MW; F807404B79F373F9 CRC64;

Query Match 76.1%; Score 35; DB 11; Length 549;  
 Best Local Similarity 75.0%; Pred. No. 58;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 GGLCVQDG 8  
 111111  
 DB 266 GGLCVQDG 273

RESULT 13  
 Q9V4M1 PRELIMINARY; PRT; 561 AA.  
 AC 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)  
 DE CG11140 PROTEIN.  
 GN CG11140.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 CC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 CC Ephyridiida; Drosophilidae; Drosophila.  
 NCBI\_TaxID=7227;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RP STRAIN=BERKELEY;  
 RA Adams M.D., Celiker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.E.,  
 George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 Brandon R.C., Rogers Y.-H.C., Blazey R.G., Champe M., Pfeiffer B.D.,  
 Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beaskey E.M.,  
 Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 Botkova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,  
 Butts K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunfey B.C., Dunn P.,  
 Durbin K.J., Evangelista C.C., Ferraz C., Ferrelira S., Fleischmann W.,  
 Foster C., Gabrielian A.E., Gary N.S., Gelbart W.M., Glasser K.,  
 Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 Harris N.L., Harvey D., Heilmann T.J., Hernandez J.R., Houck J.,  
 Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegyan C.,  
 Jaitani M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 Laoko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pauley J.M.,  
 Palaztolo M., Platten G.S., Pan S., Pollard J., Puri V., Reshe M.G.,  
 Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,

SVirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenhach J.,  
 Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster.";  
 Science 287:2185-2195(2000).  
 DR EMBL; AE003841; AAF59247.1; -  
 DR HSSP; P11883; 1AD3.  
 DR FLYBASE; FBgn0031180; CG11140.  
 DR INTERPRO: IPR002086; -  
 DR PFM; PF00171; aldedh; 1.  
 DR PROSITE: PS00070; ALDEHYDE\_DEHYDR\_CYS; 1.  
 DR PROSITE: PS00687; ALDEHYDE\_DEHYDR\_GLU; 1.  
 SO SEQUENCE 561 AA; 62697 MW; F68ACB3F904B0D7A CRC64;

Query Match 76.1%; Score 35; DB 5; Length 561;  
 Best Local Similarity 85.7%; Pred. No. 59;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GGLCVQD 7  
 111111  
 DB 384 GGLCVND 390

RESULT 14  
 O11308 PRELIMINARY; PRT; 640 AA.  
 ID O11308  
 AC O11308;  
 DT 01-JUL-1997 (TREMBLrel. 04, Created)  
 DT 01-JUL-1997 (TREMBLrel. 04, Last sequence update)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)  
 DE SIMILAR TO VARIOLA C15L AND VACCINIA FILL.  
 GN B-M,N,L.8.  
 OS Molluscum contagiosum virus subtype 1 (MCV1).  
 CC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;  
 CC Molluscipoxvirus.  
 NCBI\_TaxID=10280;  
 RX MEDLINE=97093414; PubMed=8938976;  
 RP Martin-Gallardo A., Moratilla M., Funes J.M., Agromayor M., Nunez A.,  
 Varas A.J., Collado M., Valencia A., Lopez-Estebanz J.L.,  
 Escoban M.;  
 RT "Sequence analysis of a Molluscum contagiosum virus DNA region which  
 includes the gene encoding protein kinase 2 and other genes with  
 unique organization.";  
 RT Virus Genes 13:19-29(1996).  
 RN [12]  
 RP SEQUENCE FROM N.A.  
 RA Moratilla M., Agromayor M., Nunez A., Funes J.M., Varas A.J.,  
 Lopez-Estebanz J.L., Collado M., Pavon M., Martin-Gallardo A.;  
 RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; U86894; AAB57938.1; -  
 DR INTERPRO: IPR002221; -  
 DR PROSITE: PS00317; 4\_DISULFIDE\_CORE; UNKNOWN\_1.  
 SO SEQUENCE 640 AA; 68996 MW; 23752F0B08B49BEA CRC64;

Query Match 76.1%; Score 35; DB 14; Length 640;  
 Best Local Similarity 85.7%; Pred. No. 67;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 GGLCVQDG 8  
 111111  
 DB 151 GGLCVQDG 157



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

## OM protein - protein search, using sw model

Run on: June 13, 2001, 14:16:36 ; Search time 118.55 Seconds

(Without alignments)  
1.296 Million cell updates/sec

Title: PCT-US01-05825A-18

Perfect score: 46

Sequence: 1 GGLCVQDG 8

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 185757 seqs, 19210857 residues

Total number of hits satisfying chosen parameters: 185757

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

## Database :

Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep.\*  
2: /cgn2\_6/ptodata/2/1aa/5B.COMB.pep.\*  
3: /cgn2\_6/ptodata/2/1aa/6A.COMB.pep.\*  
4: /cgn2\_6/ptodata/2/1aa/6B.COMB.pep.\*  
5: /cgn2\_6/ptodata/2/1aa/PTUS.COMB.pep.\*  
6: /cgn2\_6/ptodata/2/1aa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	38	82.6	400	2 US-08-624-601-8	Sequence 8, Appl1
2	35	76.1	715	2 US-08-484-9938-10	Sequence 10, Appl1
3	35	76.1	715	2 US-08-484-1588-10	Sequence 10, Appl1
4	35	76.1	715	2 US-08-484-596A-10	Sequence 10, Appl1
5	35	76.1	715	2 US-08-480-150A-10	Sequence 10, Appl1
6	35	76.1	715	3 US-08-458-731-10	Sequence 10, Appl1
7	35	76.1	715	3 US-08-149-223A-10	Sequence 10, Appl1
8	35	76.1	715	2 US-08-484-9938-14	Sequence 14, Appl1
9	35	76.1	715	2 US-08-484-1588-14	Sequence 14, Appl1
10	35	76.1	715	2 US-08-484-596A-14	Sequence 14, Appl1
11	35	76.1	715	2 US-08-480-150A-14	Sequence 14, Appl1
12	35	76.1	715	3 US-08-458-731-14	Sequence 14, Appl1
13	35	76.1	715	3 US-08-149-223A-14	Sequence 14, Appl1
14	34	73.9	213	2 US-08-808-550-35	Sequence 35, Appl1
15	34	73.9	380	3 US-08-468-846-2	Sequence 2, Appl1
16	34	73.9	745	1 US-08-453-472-5	Sequence 5, Appl1
17	34	73.9	745	1 US-08-038-948-9	Sequence 9, Appl1
18	34	73.9	745	1 US-08-453-952-5	Sequence 9, Appl1
19	34	73.9	745	2 US-08-484-9938-43	Sequence 43, Appl1
20	34	73.9	745	2 US-08-862-903-5	Sequence 5, Appl1
21	34	73.9	745	2 US-08-484-1588-43	Sequence 43, Appl1
22	34	73.9	745	2 US-08-484-596A-43	Sequence 43, Appl1
23	34	73.9	745	2 US-08-480-150A-43	Sequence 43, Appl1
24	34	73.9	745	3 US-08-458-731-43	Sequence 43, Appl1
25	34	73.9	745	3 US-08-149-223A-43	Sequence 43, Appl1
26	33	71.7	115	2 US-07-903-029-4	Sequence 4, Appl1
27	33	71.7	115	2 US-07-903-029-5	Sequence 5, Appl1

28	33	71.7	142	1 US-08-681-812-2	Sequence 2, Appl1
29	33	71.7	336	1 US-08-414-926A-26	Sequence 26, Appl1
30	33	71.7	336	2 US-08-926-922-26	Sequence 26, Appl1
31	33	71.7	336	3 US-09-253-682-26	Sequence 26, Appl1
32	33	71.7	2556	1 US-08-185-432-17	Sequence 17, Appl1
33	33	71.7	2556	1 US-08-083-590A-20	Sequence 20, Appl1
34	33	71.7	2556	3 US-08-532-384-20	Sequence 20, Appl1
35	32	69.6	866	3 US-08-651-999A-1	Sequence 1, Appl1
36	32	69.6	968	3 US-08-651-999A-7	Sequence 7, Appl1
37	32	69.6	1153	4 US-08-097-997A-14	Sequence 14, Appl1
38	32	69.6	1153	4 US-08-665-574C-14	Sequence 14, Appl1
39	32	69.6	1153	4 US-08-946-994-14	Sequence 2, Appl1
40	31	68.5	655	1 US-07-736-178C-2	Sequence 2, Appl1
41	31	67.4	44	1 US-08-336-343A-17	Sequence 17, Appl1
42	31	67.4	50	1 US-08-336-343A-18	Sequence 17, Appl1
43	31	67.4	201	1 US-08-471-570-2	Sequence 2, Appl1
44	31	67.4	292	2 US-08-701-191A-40	Sequence 40, Appl1
45	31	67.4	299	2 US-08-701-191A-13	Sequence 13, Appl1

## ALIGNMENTS \*

RESULT 1  
US-08-624-601-8  
Sequence 8, Application US/08624601  
Patent No. 5882653  
GENERAL INFORMATION:  
APPLICANT: Kapet Dr., James B.  
APPLICANT: Levine Dr., Myron M.  
TITLE OF INVENTION: Vibrio cholerae O1 (CVD11) and non-O1  
TITLE OF INVENTION: CVD112 and CVD12M) serogroup vaccine strains, methods  
TITLE OF INVENTION: of making same and products thereof  
NUMBER OF SEQUENCES: 16  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Spencer & Frank  
STREET: 1100 New York Ave. N.W. Suite 300 East  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20005  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/624,601  
FILING DATE: 08-APR-1996  
CLASSIFICATION: 42A  
ATTORNEY/AGENT INFORMATION:  
NAME: Scheller Dr., John W.  
REGISTRATION NUMBER: 26,031  
REFERENCE/DOCKET NUMBER: BAWC20019P2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)414-4000  
TELEFAX: (202)414-4040  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 400 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Vibrio cholerae  
STRAIN: El tor 7946  
IMMEDIATE SOURCE:  
CLONE: zot  
US-08-624-601-8

Query Match 82.6%; Score 38; DB 2; Length 400;  
Best Local Similarity 87.5%; Pred. No. 22;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGLCVODG 8  
DB 291 GRLCVODG 298

## RESULT 2

US-08-484-993B-10  
; Sequence 10, Application US/08484993B  
; Patent No. 5837497  
; GENERAL INFORMATION:  
; APPLICANT: Harris Ph.D., Jeffrey D.  
; APPLICANT: Hsu, Kuang T.  
; APPLICANT: Podolski, Joseph S.  
; TITLE OF INVENTION: Materials and Methods for Immunocontraception  
; NUMBER OF SEQUENCES: 59  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
; STREET: 6300 Sears Tower, 233 South Wacker Drive  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: United States of America  
; ZIP: 60606-6402  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: IBM PC compatible  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/484,993B  
; FILING DATE: 09-NOV-1993  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/012,990  
; FILING DATE: 29-JAN-1993  
; APPLICATION DATA:  
; APPLICATION NUMBER: 07/973,341  
; FILING DATE: 09-NOV-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Clough, David W.  
; REGISTRATION NUMBER: 36,107  
; REFERENCE/DOCKET NUMBER: 31745  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312/474-6653  
; TELEFAX: 312/474-0448  
; TELEX: 25-3856  
; INFORMATION FOR SEQ ID NO: 10:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 715 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-484-993B-10

Query Match 76.1%; Score 35; DB 2; Length 715;  
Best Local Similarity 75.0%; Pred. No. 1.3e+02;  
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GGLCVODG 8  
DB 366 GDLCTODG 373

## RESULT 3

US-08-484-158B-10  
; Sequence 10, Application US/08484158B  
; Patent No. 5976545  
; GENERAL INFORMATION:

APPLICANT: Harris Ph.D., Jeffrey D.

APPLICANT: Hsu, Kuang T.

APPLICANT: Podolski, Joseph S.

TITLE OF INVENTION: Pharmaceutical Compositions for

IMMUNOCONTRACEPTION

NUMBER OF SEQUENCES: 61

CORRESPONDENCE ADDRESS:

ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &

BORUN

STREET: 6300 Sears Tower, 233 South Wacker Drive

CITY: Chicago

STATE: Illinois

COUNTRY: United States of America

ZIP: 60606-6402

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: IBM PC compatible

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/484,158B

FILING DATE: 07-JUNE-95

CLASSIFICATION: 514

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/149,223

FILING DATE: 09-NOV-93

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/012,990

FILING DATE: 29-JAN-93

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 07/973,341

FILING DATE: 09-NOV-92

ATTORNEY/AGENT INFORMATION:

NAME: Clough, David W.

REGISTRATION NUMBER: 36,107

REFERENCE/DOCKET NUMBER: 32794

TELECOMMUNICATION INFORMATION:

TELEPHONE: 312/474-6653

TELEFAX: 312/474-0448

TELEX: 25-3856

INFORMATION FOR SEQ ID NO: 10:

SEQUENCE CHARACTERISTICS:

LENGTH: 715 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-484-158B-10

Query Match 76.1%; Score 35; DB 2; Length 715;  
Best Local Similarity 75.0%; Pred. No. 1.3e+02;  
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GGLCVODG 8  
DB 366 GDLCTODG 373

## RESULT 4

US-08-484-596A-10  
; Sequence 10, Application US/08484596A  
; Patent No. 5981228  
; GENERAL INFORMATION:  
; APPLICANT: Harris Ph.D., Jeffrey D.  
; APPLICANT: Hsu, Kuang T.  
; APPLICANT: Podolski, Joseph S.  
; TITLE OF INVENTION: Materials and Methods for Immunocontraception  
; NUMBER OF SEQUENCES: 59  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
; STREET: 6300 Sears Tower, 233 South Wacker Drive  
; CITY: Chicago  
; STATE: Illinois



COUNTRY: United States of America  
ZIP: 60606-6402  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/484,596A  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/149,223  
FILING DATE: 11-NOV-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/973,341  
FILING DATE: 09-NOV-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Clough, David W.  
REGISTRATION NUMBER: 36,107  
REFERENCE/DOCKET NUMBER: 31745  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312/474-6653  
TELEFAX: 312/474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 715 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-484-596A-10

Query Match 76.1%; Score 35; DB 2; Length 715;  
Best Local Similarity 75.0%; Pred. No. 1.3e+02;  
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GGLCVQDG 8  
DB 366 GDLCTQDG 373

RESULT 5  
US-08-480-150A-10  
Sequence 10, Application US/08480150A  
Patent No. 5989550  
GENERAL INFORMATION:  
APPLICANT: Harris Ph.D., Jeffrey D.  
APPLICANT: Hsu, Kuang T.  
APPLICANT: Podolski, Joseph S.  
TITLE OF INVENTION: Materials and Methods for Immunocontraception  
NUMBER OF SEQUENCES: 59  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
STREET: 6300 Sears Tower, 233 South Wacker Drive  
CITY: Chicago  
STATE: Illinois  
COUNTRY: United States of America  
ZIP: 60606-6402  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/480,150A  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/149,223  
FILING DATE: 09-NOV-1993  
APPLICATION NUMBER: 08/012,990

FILING DATE: 29-JAN-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/973,341  
FILING DATE: 09-NOV-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Clough, David W.  
REGISTRATION NUMBER: 36,107  
REFERENCE/DOCKET NUMBER: 31745  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312/474-6653  
TELEFAX: 312/474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 715 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-480-150A-10

Query Match 76.1%; Score 35; DB 2; Length 715;  
Best Local Similarity 75.0%; Pred. No. 1.3e+02;  
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GGLCVQDG 8  
DB 366 GDLCTQDG 373

RESULT 6  
US-08-458-731-10  
Sequence 10, Application US/08458731  
Patent No. 6001599  
GENERAL INFORMATION:  
APPLICANT: Harris Ph.D., Jeffrey D.  
APPLICANT: Hsu, Kuang T.  
APPLICANT: Podolski, Joseph S.  
TITLE OF INVENTION: Materials and Methods for Immunocontraception  
NUMBER OF SEQUENCES: 59  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
STREET: 6300 Sears Tower, 233 South Wacker Drive  
CITY: Chicago  
STATE: Illinois  
COUNTRY: United States of America  
ZIP: 60606-6402  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/458,731  
FILING DATE: 09-NOV-1993  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/012,990  
FILING DATE: 29-JAN-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/973,341  
FILING DATE: 09-NOV-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Clough, David W.  
REGISTRATION NUMBER: 36,107  
REFERENCE/DOCKET NUMBER: 31745  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312/474-6653  
TELEFAX: 312/474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 715 amino acids

TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-458-731-10

Query Match 76.1%; Score 35; DB 3; Length 715;  
Best Local Similarity 75.0%; Pred. No. 1.3e+02;  
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GGLCVQDG 8  
111111  
DB 366 GDLCTQDG 373

RESULT 7  
US-08-149-223A-10  
Sequence 10, Application US/08149223A  
Patent No. 6027727  
GENERAL INFORMATION:  
APPLICANT: Harris Ph.D., Jeffrey D.  
APPLICANT: Hsu, Kuang T.  
APPLICANT: Podolski, Joseph S.  
TITLE OF INVENTION: Materials and Methods for Immunococontraception  
NUMBER OF SEQUENCES: 59  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
STREET: 6300 Sears Tower, 233 South Wacker Drive  
CITY: Chicago  
STATE: Illinois  
COUNTRY: United States of America  
ZIP: 60606-6402  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/149/223A  
FILING DATE: 09-NOV-1993  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/012,990  
FILING DATE: 29-JAN-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/973,341  
FILING DATE: 09-NOV-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Clough, David W.  
REGISTRATION NUMBER: 36,107  
REFERENCE/DOCKET NUMBER: 31745  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312/474-6653  
TELEFAX: 312/474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 715 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-149-223A-10

Query Match 76.1%; Score 35; DB 3; Length 715;  
Best Local Similarity 75.0%; Pred. No. 1.3e+02;  
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GGLCVQDG 8  
111111  
DB 366 GDLCTQDG 373

RESULT 8  
US-08-484-993B-14  
Sequence 14, Application US/08484993B  
Patent No. 5837497  
GENERAL INFORMATION:  
APPLICANT: Harris Ph.D., Jeffrey D.  
APPLICANT: Hsu, Kuang T.  
APPLICANT: Podolski, Joseph S.  
TITLE OF INVENTION: Materials and Methods for Immunococontraception  
NUMBER OF SEQUENCES: 59  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
STREET: 6300 Sears Tower, 233 South Wacker Drive  
CITY: Chicago  
STATE: Illinois  
COUNTRY: United States of America  
ZIP: 60606-6402  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/484,993B  
FILING DATE: 09-NOV-1993  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/012,990  
FILING DATE: 29-JAN-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/973,341  
FILING DATE: 09-NOV-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Clough, David W.  
REGISTRATION NUMBER: 36,107  
REFERENCE/DOCKET NUMBER: 31745  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312/474-6653  
TELEFAX: 312/474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 716 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-484-993B-14

Query Match 76.1%; Score 35; DB 2; Length 716;  
Best Local Similarity 75.0%; Pred. No. 1.3e+02;  
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GGLCVQDG 8  
111111  
DB 368 GDLCTQDG 375

RESULT 9  
US-08-484-158B-14  
Sequence 14, Application US/08484158B  
Patent No. 5976545  
GENERAL INFORMATION:  
APPLICANT: Harris Ph.D., Jeffrey D.  
APPLICANT: Hsu, Kuang T.  
APPLICANT: Podolski, Joseph S.  
TITLE OF INVENTION: Pharmaceutical Compositions for  
TITLE OF INVENTION: Immunococontraception  
NUMBER OF SEQUENCES: 61  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
STREET: 6300 Sears Tower, 233 South Wacker Drive

CITY: Chicago  
STATE: Illinois  
COUNTRY: United States of America  
ZIP: 60606-6402  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/484,1588  
FILING DATE: 07-JUNE-95  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/149,223  
FILING DATE: 09-NOV-93  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/012,990  
FILING DATE: 29-JAN-93  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/973,341  
FILING DATE: 09-NOV-92  
ATTORNEY/AGENT INFORMATION:  
NAME: Clough, David W.  
REGISTRATION NUMBER: 36,107  
REFERENCE/DOCKET NUMBER: 32794  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312/474-6653  
TELEFAX: 312/474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 716 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-484-1588-14

Query Match 76.1%; Score 35; DB 2; Length 716;  
Best Local Similarity 75.0%; Pred. NO. 1.3e+02;  
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 GGLCVQDG 8  
| | | | |  
DB 368 GDLCTQDG 375

RESULT 10  
US-08-484-596A-14  
Sequence 14, Application US/08484596A  
Patent No. 5981228  
GENERAL INFORMATION:  
APPLICANT: Harris Ph.D., Jeffrey D.  
APPLICANT: Hsu, Kuang T.  
APPLICANT: Podolski, Joseph S.  
TITLE OF INVENTION: Materials and Methods for Immunocontraception  
NUMBER OF SEQUENCES: 59  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
STREET: 6300 Sears Tower, 233 South Wacker Drive  
CITY: Chicago  
STATE: Illinois  
COUNTRY: United States of America  
ZIP: 60606-6402  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/484,596A  
FILING DATE:

CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/149,223  
FILING DATE: 11-NOV-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/973,341  
FILING DATE: 09-NOV-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Clough, David W.  
REGISTRATION NUMBER: 36,107  
REFERENCE/DOCKET NUMBER: 31745  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312/474-6653  
TELEFAX: 312/474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 716 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-484-596A-14

Query Match 76.1%; Score 35; DB 2; Length 716;  
Best Local Similarity 75.0%; Pred. NO. 1.3e+02;  
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 GGLCVQDG 8  
| | | | |  
DB 368 GDLCTQDG 375

RESULT 11  
US-08-480-150A-14  
Sequence 14, Application US/08480150A  
Patent No. 5989550  
GENERAL INFORMATION:  
APPLICANT: Harris Ph.D., Jeffrey D.  
APPLICANT: Hsu, Kuang T.  
APPLICANT: Podolski, Joseph S.  
TITLE OF INVENTION: Materials and Methods for Immunocontraception  
NUMBER OF SEQUENCES: 59  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
STREET: 6300 Sears Tower, 233 South Wacker Drive  
CITY: Chicago  
STATE: Illinois  
COUNTRY: United States of America  
ZIP: 60606-6402  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/480,150A  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/149,223  
FILING DATE: 09-NOV-1993  
APPLICATION NUMBER: 08/012,990  
FILING DATE: 29-JAN-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/973,341  
FILING DATE: 09-NOV-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Clough, David W.  
REGISTRATION NUMBER: 36,107  
REFERENCE/DOCKET NUMBER: 31745  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312/474-6653

TELEFAX: 312/474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 716 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-480-150A-14

Query Match 76.1%; Score 35; DB 2; Length 716;  
Best Local Similarity 75.0%; Pred. No. 1.3e+02;  
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GGLCYODG 8  
DB 368 GDLCYODG 375

RESULT 12  
US-08-458-731-14  
Sequence 14, Application US/08458731  
Patent No. 6001399  
GENERAL INFORMATION:  
APPLICANT: Harris Ph.D., Jeffrey D.  
APPLICANT: Hsu, Kuang T.  
APPLICANT: Podolski, Joseph S.  
TITLE OF INVENTION: Materials and Methods for Immunocontraception  
NUMBER OF SEQUENCES: 59  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
STREET: 6300 Sears Tower, 233 South Wacker Drive  
CITY: Chicago  
STATE: Illinois  
COUNTRY: United States of America  
ZIP: 60606-6402  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: IBM PC compatible  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/458,731  
FILING DATE: 09-NOV-1993  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/012,990  
FILING DATE: 29-JAN-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/973,341  
FILING DATE: 09-NOV-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Clough, David W.  
REGISTRATION NUMBER: 36,107  
REFERENCE/DOCKET NUMBER: 31745  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312/474-6653  
TELEFAX: 312/474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 716 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-458-731-14

Query Match 76.1%; Score 35; DB 3; Length 716;  
Best Local Similarity 75.0%; Pred. No. 1.3e+02;  
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GGLCYODG 8  
DB 368 GDLCYODG 375

RESULT 13  
US-08-149-223A-14  
Sequence 14, Application US/08149223A  
Patent No. 6027727  
GENERAL INFORMATION:  
APPLICANT: Harris Ph.D., Jeffrey D.  
APPLICANT: Hsu, Kuang T.  
APPLICANT: Podolski, Joseph S.  
TITLE OF INVENTION: Materials and Methods for Immunocontraception  
NUMBER OF SEQUENCES: 59  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
STREET: 6300 Sears Tower, 233 South Wacker Drive  
CITY: Chicago  
STATE: Illinois  
COUNTRY: United States of America  
ZIP: 60606-6402  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: IBM PC compatible  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/149,223A  
FILING DATE: 09-NOV-1993  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/012,990  
FILING DATE: 29-JAN-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/973,341  
FILING DATE: 09-NOV-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Clough, David W.  
REGISTRATION NUMBER: 36,107  
REFERENCE/DOCKET NUMBER: 31745  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312/474-6653  
TELEFAX: 312/474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 716 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-149-223A-14

Query Match 76.1%; Score 35; DB 3; Length 716;  
Best Local Similarity 75.0%; Pred. No. 1.3e+02;  
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GGLCYODG 8  
DB 368 GDLCYODG 375

RESULT 14  
US-08-808-550-35  
Sequence 35, Application US/08808550  
Patent No. 5871992  
GENERAL INFORMATION:  
APPLICANT: Teebor, George W.  
APPLICANT: Hilbert, Timothy P.  
TITLE OF INVENTION: MAMMALIAN ENDOCLINASE III AND  
DIAGNOSTIC AND THERAPEUTIC USES THEREOF  
NUMBER OF SEQUENCES: 42

```

CORRESPONDENCE ADDRESS:
ADDRESSSEE: David A. Jackson, Esq.
STREET: 411 Hackensack Ave, Continental Plaza, 4th
STREET: Floor
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601

COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA: US/08/808, 550
APPLICATION NUMBER: US/08/808, 550
FILING DATE: 26-FEB-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 1049-1-001 N
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-487-5800
TELEFAX: 201-343-1684
INFORMATION FOR SEQ ID NO: 35:
SEQUENCE CHARACTERISTICS:
LENGTH: 213 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
FRAGMENT TYPE: N-terminal
ORIGINAL SOURCE:
ORGANISM: S. pombe
US-08-808-550-35

Query Match      73.9%; Score 34; DB 2; Length 213;
Best Local Similarity 71.4%; Pred. No. 56;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGLCYOD 7
    |||::||
DB 60 GGLCLED 66

RESULT 15
US-08-468-846-2
: Sequence 2, Application US/08468846
: Patent No. 6074839
: GENERAL INFORMATION:
: APPLICANT: Melsner, Paul
: APPLICANT: Fuldner, Rebecca
: APPLICANT: Fei-Wel, Ying
: APPLICANT: Adams, Mark
: TITLE OF INVENTION: TRANSFORMING GROWTH FACTOR ALPHA HI
: NUMBER OF SEQUENCES: 15
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN, CECCHI,
: ADDRESSEE: STUART & OLSTEIN
: STREET: 6 Becker Farm Road
: CITY: Roseland
: STATE: New Jersey
: COUNTRY: USA
: ZIP: 07068
: COMPUTER READABLE FORM:
: MEDIUM TYPE: floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/468, 846

```

```

FILING DATE: 06-JUN-1995
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/208,008
FILING DATE: 08-MAR-1994
ATTORNEY/AGENT INFORMATION:
NAME: Parraro, Gregory D.
REGISTRATION NUMBER: 36,134
REFERENCE/DOCKET NUMBER: 325800-465
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-994-1700
TELEFAX: 201-994-1744
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 380 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-468-846-2

Query Match      73.9%; Score 34; DB 3; Length 380;
Best Local Similarity 62.5%; Pred. No. 99;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGLCYODG 8
    ||:|:|
DB 85 GGVCKEDG 92

```

Search completed: June 13, 2001, 14:16:37  
Job time: 498 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

# OM protein - protein search, using sw model

Run on: June 13, 2001, 14:14:39 ; Search time 229.28 Seconds  
(without alignments)  
1.995 Million cell updates/sec

Title: PCT-US01-05825A-19

Perfect score: 42

Sequence: 1 GGLVQPG 8

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 390729 seqs, 57163235 residues

Total number of hits satisfying chosen parameters: 390729

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A\_Geneseq\_0401.\*  
1: /SIDS6/gcgdata/geneseq/geneseq/AA1980.DAT.\*  
2: /SIDS6/gcgdata/geneseq/geneseq/AA1981.DAT.\*  
3: /SIDS6/gcgdata/geneseq/geneseq/AA1982.DAT.\*  
4: /SIDS6/gcgdata/geneseq/geneseq/AA1983.DAT.\*  
5: /SIDS6/gcgdata/geneseq/geneseq/AA1984.DAT.\*  
6: /SIDS6/gcgdata/geneseq/geneseq/AA1985.DAT.\*  
7: /SIDS6/gcgdata/geneseq/geneseq/AA1986.DAT.\*  
8: /SIDS6/gcgdata/geneseq/geneseq/AA1987.DAT.\*  
9: /SIDS6/gcgdata/geneseq/geneseq/AA1988.DAT.\*  
10: /SIDS6/gcgdata/geneseq/geneseq/AA1989.DAT.\*  
11: /SIDS6/gcgdata/geneseq/geneseq/AA1990.DAT.\*  
12: /SIDS6/gcgdata/geneseq/geneseq/AA1991.DAT.\*  
13: /SIDS6/gcgdata/geneseq/geneseq/AA1992.DAT.\*  
14: /SIDS6/gcgdata/geneseq/geneseq/AA1993.DAT.\*  
15: /SIDS6/gcgdata/geneseq/geneseq/AA1994.DAT.\*  
16: /SIDS6/gcgdata/geneseq/geneseq/AA1995.DAT.\*  
17: /SIDS6/gcgdata/geneseq/geneseq/AA1996.DAT.\*  
18: /SIDS6/gcgdata/geneseq/geneseq/AA1997.DAT.\*  
19: /SIDS6/gcgdata/geneseq/geneseq/AA1998.DAT.\*  
20: /SIDS6/gcgdata/geneseq/geneseq/AA1999.DAT.\*  
21: /SIDS6/gcgdata/geneseq/geneseq/AA2000.DAT.\*  
22: /SIDS6/gcgdata/geneseq/geneseq/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	42	100.0	8	21	Y79123
2	39	92.9	8	21	Y84661
3	39	92.9	8	21	Y79119
4	39	92.9	20	20	W94489
5	39	92.9	20	21	Y79133
6	39	92.9	25	15	R52775
7	39	92.9	25	15	R52793
8	38	90.5	498	21	B07577
9	37	88.1	8	21	Y79121
10	37	88.1	18	19	W68167
11	37	88.1	20	21	Y84660

12	37	88.1	118	16	R66314	Human immunoglobulin
13	37	88.1	120	17	W00240	EGF receptor chime
14	37	88.1	121	15	B53648	Human colon cancer
15	37	88.1	123	21	R54811	SpA-reactive VH re
16	36	85.7	98	21	B40074	Anti-hiL12 antibod
17	36	85.7	118	16	R66331	Human immunoglobul
18	36	85.7	124	19	W80816	Amino acid sequenc
19	35	83.3	115	21	Y96064	Human anti-DAP ant
20	35	83.3	118	20	Y15401	Protein encoded by
21	35	83.3	119	19	W69322	15D3 antibody heav
22	35	83.3	119	20	Y32833	Antibody 15D3 heav
23	35	83.3	125	20	W73503	Anti-hcg anti-R6
24	35	83.3	225	20	Y15409	Anti-hcg anti-R6
25	35	83.3	239	20	Y15408	Anti-hcg anti-R6
26	35	83.3	500	21	B07654	Llama antibody fra
27	34	81.0	8	21	Y79111	Peptide antagonist
28	34	81.0	8	21	Y79117	Peptide antagonist
29	34	81.0	8	21	Y79127	Peptide antagonist
30	34	81.0	8	21	Y79127	Peptide antagonist
31	34	81.0	13	20	W86096	Peptide from human
32	34	81.0	13	20	W86076	Peptide from human
33	34	81.0	13	20	W86088	Peptide from human
34	34	81.0	18	16	W73953	IGG heavy chain fr
35	34	81.0	18	16	R82833	N-terminal of C179
36	34	81.0	20	19	W68169	Human Igg heavy ch
37	34	81.0	20	20	W94487	Human adult heart
38	34	81.0	20	21	Y79130	Human adult heart
39	34	81.0	27	18	W16581	Anti-RSV F glycopr
40	34	81.0	30	17	R87049	Human group III he
41	34	81.0	30	21	Y68811	Human heavy chain
42	34	81.0	37	17	W00242	EGF receptor chime
43	34	81.0	40	7	P61028	H-chain variable r
44	34	81.0	62	16	R76369	HSV-neutralising a
45	34	81.0	73	21	Y64774	Human 5' EST relat

## ALIGNMENTS

RESULT	1
ID	Y79123
Y79123	standard; Peptide; 8 AA.
XX	
AC	Y79123;
XX	
DT	05-JUN-2000 (first entry)
XX	
DE	Peptide antagonist of zonulin.
XX	
KW	Zonulin; antagonist; zonula occludens toxin receptor;
KW	blood-brain barrier; antiinflammatory; cerebroprotective;
KW	neuroprotective; dermatological; antitumor; antiviral;
KW	antibacterial; cytostatic; anti-HIV; vulnereary; antiallergic;
KW	hypotensive; immunosuppressive; antiparasitic; vasotropic;
XX	gastrointestinal inflammation; therapy.
XX	
OS	Synthetic.
XX	
PN	W0200007609-A1.
XX	
PD	17-FEB-2000.
XX	
PF	28-JUL-1999; 99WO-US16683.
XX	
PR	03-AUG-1998; 98US-0127815.
XX	
PA	(UYMA-) UNIV MARYLAND BALTIMORE.
XX	
PI	Murine BrE-3 immun
XX	
XX	Protein encoded by
DR	Peptide antagonist
XX	Complex scupa/supa
PT	N-terminal sequenc

New peptide antagonist of zonulin useful as antiinflammatory agent for

PT threatening cerebral ischemia, stroke, cerebral edema, gastritis,  
PT shigellosis, viral gastroenteritis, meningitis, encephalomyelitis  
XX  
PS Claim 1, Page 47; 69pp; English.

CC This present sequence is that of a peptide antagonist of zonulin  
CC (2), one of 25 such peptides (see Y79105-29) of the invention,  
CC which bind to a zonula occludens toxin (ZOT) receptor, yet do not  
CC physiologically modulate the opening of mammalian tight junctions  
CC (TJ). The peptide antagonists are based on a common motif of ZOT  
CC and human zonulins, which is believed to be critical for receptor  
CC binding. They can be prepared by chemical synthesis or by use of  
CC recombinant DNA techniques. The peptide antagonists are used as an  
CC antiinflammatory agents in the treatment of gastrointestinal  
CC inflammation, where they bind to the ZOT receptor in the intestine  
CC and yet does not physiologically modulate the opening of TJ in the  
CC intestine. Gastrointestinal inflammation conditions give rise to  
CC increased intestinal permeability and the peptide is useful for  
CC treating intestinal conditions that cause protein losing enteropathy  
CC caused by infection, e.g. Clostridium difficile infection,  
CC enterocolitis, shigellosis, viral gastroenteritis, parasite  
CC infestation, bacterial overgrowth, whipple's disease, diseases with  
CC mucosal erosion or ulcerations, e.g. gastritis, gastric cancer,  
CC collagenous colitis, inflammatory bowel disease, diseases marked by  
CC lymphatic obstruction, e.g. congenital intestinal lymphangiectasia,  
CC sarcoidosis lymphoma, mesenteric tuberculosis, and after surgical  
CC correction of congenital heart disease with Fontan's operation,  
CC mucosal diseases without ulceration, e.g. Menetrier's disease,  
CC coliac disease, eosinophilic gastroenteritis, and immune diseases,  
CC e.g. systemic lupus erythematosus or food allergies, primarily to  
CC milk.

**SQ Sequence 8 AA;**

Query Match	100.0%	Score 42	. DB 21	Length 8
Best Local Similarity	100.0%	Pred. No.	3.2e+05	
Matches	8	Conservative	0	Mismatches 0
				Indels 0
				Gaps 0

QY	1	GGLVQPG	8
Db	1	gglLvqpg	8

RESULT	2
Y84661	
ID	Y84661 standard; Protein; 8 AA.

AC Y84661;

DT 25-JUL-2000 (first entry)

DE Peptide antagonist FZI/0 of zonula occludens toxin (zot) polypeptide.

KM Human, zot; zonula occludens toxin; zonulin; antigen presenting cell;  
 KM APC; lymphocyte proliferation; antigen; auto-immune disorder;  
 KM Immune-related disorder; immune system rejection; multiple sclerosis;  
 KM organ transplantation; inflammatory disease; allergic disease;  
 KM rheumatoid arthritis; insulin dependent diabetes mellitus;  
 KM celiac disease; Sjogren's syndrome; systemic lupus erythematosus;  
 KM auto-immune thyroiditis; idiopathic thrombocytopenic purpura;  
 KM hemolytic anemia; Grave's disease; Addison disease; autoimmune orchitis;  
 KM pernicious anemia; vasculitis; autoimmune coagulopathy; polyomyelitis;  
 KM metasthema gravis; polyneuritis; pemphigus; rheumatic carditis;  
 KM dermatomyositis; scleroderma; asthma; psoriasis; eczematous dermatitis;  
 KM Kaposi's sarcoma; inflammatory bowel disease; proliferative disorder;  
 KM PCR primer; ss.

OS Synthetic.

PN WO200015252-A1

PD 23-MAR-2000

XX	09-SEP-1999;	99WO-US18842
PF		
XX		
PR	14-SEP-1998;	98US-0100266

PT Suppression of antigen presenting cell mediated lymphocyte  
PT proliferation, by administering a zot-related immunoregulator usefull  
PT for treating immune-related disorders, immune system rejection  
PT . subsequent to tissue or organ transplantation -  
XX  
PS Example 7; Page 59; 95pp; English.

Example 7; Page 59; 95pp; English

The present sequence represents a peptide antagonist of zonula occludens toxin (zot) polypeptide. The specification describes a method of suppressing antigen presenting cell (APC)-mediated lymphocyte proliferation in a mammalian host pre-exposed to a particular antigen. The method comprises administering to the host an effective amount of a zot-related immunoregulator selected from zot (zonula occludens toxin) or zonulin, the amount effective to down-regulate the activity of the APC. The method can be used to down-regulate APC-mediated lymphocyte proliferation in mammalian hosts suffering from auto-immune or immune-related disorders, immune system rejection subsequent to tissue or organ transplantation, or inflammatory or allergic diseases. The auto-immune or immune related disorders include multiple sclerosis, rheumatoid arthritis, insulin dependent diabetes mellitus, celiac disease, Sjogren's syndrome, systemic lupus erythematosus, auto-immune thyroiditis, idiopathic thrombocytopenic purpura, hemolytic anemia, Grave's disease, Addison disease, autoimmune orchitis, pernicious anemia, vasculitis, autoimmune coagulopathies, myasthenia gravis, polyneuritis, pemphigus, rheumatic carditis, polymyositis, dermatomyositis, and scleroderma. The inflammatory or allergic disease or disorder is selected from asthma, psoriasis, eczematous dermatitis, Kaposi's sarcoma, multiple sclerosis, inflammatory bowel disease, proliferative disorders of smooth muscle cells, and inflammatory conditions associated with mycotic, viral, parasitic, or bacterial infections.

Sequence 8 AA;

Query Match	92.9%	Score 39;	DB 21;	length 8;
Best Local Similarity	87.5%	Pred. No.	3.2e+05;	
Matches	7;	Conservative	1;	Mismatches 0;
				Gaps 0

QY	1	GGLVQPG	8
		:	
Db	1	gvlvqpg	8

RESULT	3
Y79119	
ID	Y79119 standard; Peptide; 8 AA

AC Y79119;

DT 05-JUN-2000 (first entry)

Peptide antagonist of zonulin.

KM zonulin; antagonist; zonula occludens toxin receptor; human;  
blood-brain barrier; antiinflammatory; cerebroprotective;  
KM neuroprotective; dermatological; anticancer; antiviral;  
KM antihisterial; cytostatic; anti-HIV; vulnereary; antiallergic  
KM hypotensive; immunosuppressive; antiparasitic; vasotropic;  
KM gastrointestinal inflammation; therapy.

OS Homo sapiens.







CC Primers J02, J03, J04, J014 and VH1BACK (062765-062769) were all  
CC used to prepare cDNAs that encode the BrE-3 mouse Ig variable  
CC domains. The amplified V-regions lacked constant regions so as to  
CC produce less immunogenic polypeptides. A hybrid polypeptide was  
CC prepared using human constant regions with the murine V regions.  
CC The chimeric polypeptide retained the binding affinity of BrE-3  
CC for human milk fat globule. The amino acid sequences deduced from  
CC the amplified VI and VH regions were compared to N-terminal  
CC sequences which had been determined directly (i.e. R52792 and  
CC R52793). The general agreement between the predicted and the  
CC determined amino acid sequences that the cloned cDNAs encode BrE-3.  
CC  
CC  
CC  
SQ Sequence 25 AA:

Query Match 92.9%; Score 39; DB 15; Length 25;  
Best Local Similarity 87.5%; Pred. No. 0.83;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GGLVQPG 8  
||:|||||  
Db 8 ggvlvqpg 15

RESULT 8  
B07577  
ID B07577 standard; Protein: 498 AA.  
XX B07577;  
AC  
XX  
XX 20-OCT-2000 (first entry)  
DT  
XX  
XX Protein encoded by the bleomycin (BLM) gene cluster ORF30.  
DE  
XX  
XX BLM gene cluster; bleomycin gene cluster; polyketide metabolite;  
KW bleomycin; bleomycin analogue; holo-carrier protein; thiazolidine;  
KW thiazolidine; bithiazolidine; microbial metabolite; sugar.  
XX  
XX Streptomyces verticillius.  
OS  
XX  
XX Key Location/Qualifiers  
FH Misc-difference 1  
FT /note= "encoded by GNG"  
FT  
XX  
XX WO200040704-A1.  
PN  
XX  
XX 13-JUL-2000.  
PD  
XX  
XX 06-JAN-2000; 2000WO-US00445.  
PF  
XX  
XX 06-JAN-1999; 99US-0115435.  
PR 05-FEB-1999; 99US-0118848.  
PR 05-JAN-2000; 2000US-0477962.  
XX  
XX (RECC ) UNIV CALIFORNIA.  
PA  
XX  
XX Shen B, Du L, Sanchez C, Chen M, Edwards DJ;  
PI  
XX  
XX WPI; 2000-465974/40.  
DR N-PSDB; A58471.  
DR  
XX  
XX New bleomycin gene cluster components useful for peptide and/or  
PT polyketide metabolites, especially bleomycin, production and for  
PT chemically modifying biological molecules -  
PS  
XX  
XX Disclosure; Page 97-137; 162pp; English.  
XX  
XX B07556-78 represent proteins encoded by open reading frames (ORFs)  
CC 8 to 30 of the BLM (Bleomycin) gene cluster. The proteins encoded  
CC by the gene cluster are useful for producing peptides and/or polyketide  
CC metabolites, especially bleomycin or bleomycin analogues. They are  
CC also useful for chemically modifying biological molecules to produce  
CC branched methyl groups, and for coupling amino acids and fatty

CC acids. They may be reacted with an apo-carrier protein and coenzyme A  
CC to produce a holo-carrier protein. The BLM gene cluster or catalytic  
CC domains can be used individually or collectively to produce  
CC thiazolidine, thiazoline, bithiazoline and bithiazoline-containing  
CC microbial metabolites. The BLM gene cluster may also be used to produce  
CC sugars.  
CC  
CC  
SQ Sequence 498 AA:

Query Match 90.5%; Score 38; DB 21; Length 498;  
Best Local Similarity 87.5%; Pred. No. 27;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GGLVQPG 8  
|||:|||||  
Db 98 gglfvqpg 105

RESULT 9  
Y79121  
ID Y79121 standard; Peptide: 8 AA.  
XX  
XX Y79121;  
AC  
XX  
XX 05-JUN-2000 (first entry)  
DT  
XX  
XX Peptide antagonist of zonulin.  
DE  
XX  
XX Zonulin; antagonist; zonula occludens toxin receptor;  
KW blood-brain barrier; antiinflammatory; cerebroprotective;  
KW neuroprotective; dermatological; antitumor; antiviral;  
KW antibacterial; cytostatic; anti-HIV; vulnerability; antiallergic;  
KW hypotensive; immunosuppressive; antiparasitic; vasotropic;  
KW gastrointestinal inflammation; therapy.  
KW  
XX  
XX Synthetic.  
OS  
XX  
XX WO200007609-A1.  
PN  
XX  
XX 17-FEB-2000.  
PD  
XX  
XX 28-JUL-1999; 99WO-US16683.  
PF  
XX  
XX 03-AUG-1998; 98US-0127815.  
PR  
XX  
XX (UYMA-) UNIV MARYLAND BALTIMORE.  
PA  
XX  
XX Fasano A;  
PI  
XX  
XX WPI; 2000-205565/18.  
DR  
XX  
XX New peptide antagonist of zonulin useful as antiinflammatory agent for  
PT treating cerebral ischemia, stroke, cerebral edema, gastritis,  
PT shigellosis, viral gastroenteritis, meningitis, encephalomyelitis -  
PT  
XX  
XX Claim 1; Page 46; 69pp; English.  
PS  
XX  
XX This present sequence is that of a peptide antagonist of zonulin  
CC (Z), one of 25 such peptides (see Y79105-29) of the invention,  
CC which bind to a zonula occludens toxin (ZOT) receptor. Yet do not  
CC physiologically modulate the opening of mammalian tight junctions  
CC (TJ). The peptide antagonists are based on a common motif of ZOT  
CC and human zonulins, which is believed to be critical for receptor  
CC binding. They can be prepared by chemical synthesis or by use of  
CC recombinant DNA techniques. The peptide antagonists are used as an  
CC antiinflammatory agents in the treatment of gastrointestinal  
CC inflammation, where they bind to the ZOT receptor in the intestine  
CC and yet does not physiologically modulate the opening of TJ in the  
CC intestine. Gastrointestinal inflammation conditions give rise to  
CC increased intestinal permeability and the peptide is useful for  
CC treating intestinal conditions that cause protein losing enteropathy  
CC caused by infection, e.g. Clostridium difficile infection.

CC enterocolitis, shigellosis, viral gastroenteritis, parasite  
 CC infestation, bacterial overgrowth, whipple's disease, diseases with  
 CC mucosal erosion or ulcerations, e.g. gastritis, gastric cancer,  
 CC collagenous colitis, inflammatory bowel disease, diseases marked by  
 CC lymphatic obstruction, e.g. congenital intestinal lymphangiectasia,  
 CC sarcoidosis lymphoma, mesenteric tuberculosis, and after surgical  
 CC correction of congenital heart disease with Fontan's operation,  
 CC mucosal diseases without ulceration, e.g. Meckler's disease,  
 CC coliac disease, eosinophilic gastroenteritis, and immune diseases,  
 CC e.g. systemic lupus erythematosus or food allergies, primarily to  
 CC milk.  
 SQ Sequence 8 AA;  
 XX  
 XX  
 Query Match 88.1%; Score 37; DB 21; Length 8;  
 Best Local Similarity 87.5%; Pred. No. 3.2e+05;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 GGLVQPG 8  
 ||| |||  
 Db 1 gglcvpg 8  
 RESULT 10  
 W68167  
 ID W68167 standard; peptide; 18 AA.  
 XX  
 AC W68167;  
 XX  
 DT 06-OCT-1998 (first entry)  
 XX  
 DE Complex scuPA/suPAR fibrinolytic activity regulating peptide 2.  
 XX  
 KW Thrombolytic activity: soluble urokinase plasminogen activator receptor;  
 KW single chain urokinase type plasminogen activator; thromboembolism;  
 KW scuPA; suPAR; IgG; regulator; fibrinolytic activity; fibrin clot;  
 KW myocardial infarction; cerebro-vascular event; pulmonary embolism;  
 KW deep vein thrombosis; immunoglobulin; human.  
 XX  
 OS Synthetic.  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT MISC-difference 10 /note="unidentified"  
 FT  
 XX  
 PN MO9825641-A1.  
 XX  
 PD 18-JUN-1998.  
 XX  
 PF 09-DEC-1997; 97MO-IL00402.  
 XX  
 PR 09-DEC-1996; 96US-0032676.  
 XX  
 PA (HADA-) HADASIT MEDICAL RES SERVICES & DEV.  
 XX  
 PI Hlgaz1 AA;  
 XX  
 DR WPI; 1998-348262/30.  
 XX  
 PT New compositions with thrombolytic activity for, e.g. treatment of  
 PT thromboembolism - comprise complex of single chain urokinase type  
 PT plasminogen activator, scuPA, and soluble urokinase plasminogen  
 PT activator receptor, suPAR  
 XX  
 PS Example 5; Page 24; 46pp; English.  
 CC This represents a human immunoglobulin G (IgG) derived peptide sequence  
 CC that has a stimulating effect on the fibrinolytic activity of the scuPA/  
 CC suPAR complex of the invention. This sequence has similarity to the human  
 CC Ig heavy chain, IG V-III region (HV37). The invention provides a  
 CC thrombolytic therapeutic composition for the treatment and prevention of

CC a thromboembolic disorder associated with the formation of fibrin clots.  
 CC The composition comprises, as the active ingredient, a complex of a  
 CC single chain urokinase type plasminogen activator (scuPA) and a soluble  
 CC urokinase plasminogen activator receptor (suPAR). The complex (scuPA/  
 CC suPAR) has thrombolytic activity under physiological conditions and in  
 CC the presence of IgG, or of at least 1 IgG-derived peptide, and induces  
 CC fibrinolysis of fibrin clots. The compositions and complex are useful  
 CC for the treatment or prevention of thromboembolic disorders associated  
 CC with the formation of fibrin clots, especially myocardial infarctions,  
 CC cerebro-vascular events, pulmonary embolism and deep vein thrombosis.  
 XX  
 SQ Sequence 18 AA;  
 XX  
 XX  
 Query Match 88.1%; Score 37; DB 19; Length 18;  
 Best Local Similarity 87.5%; Pred. No. 1.4;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 GGLVQPG 8  
 ||| |||  
 Db 8 gxxlvpg 15  
 RESULT 11  
 Y84660  
 ID Y84660 standard; peptide; 20 AA.  
 XX  
 AC Y84660;  
 XX  
 DT 25-JUL-2000 (first entry)  
 XX  
 DE N-terminal sequence of a human zonulin protein of 47 kDa.  
 XX  
 KW Human; Zot; zonula occludens toxin; zonulin; antigen presenting cell;  
 KW APC; lymphocyte proliferation; antigen; auto-immune disorder;  
 KW immune-related disorder; immune system rejection; multiple sclerosis;  
 KW organ transplantation; inflammatory disease; allergic disease;  
 KW rheumatoid arthritis; insulin dependent diabetes mellitus;  
 KW celiac disease; Sjogren's syndrome; systemic lupus erythematosus;  
 KW auto-immune thyroiditis; idiopathic thrombocytopenic purpura;  
 KW hemolytic anemia; Grave's disease; Addison disease; autoimmune orchitis;  
 KW pernicious anemia; vasculitis; autoimmune coagulopathy; polyarthritis;  
 KW myasthenia gravis; polyneuritis; pemphigus; rheumatic carditis;  
 KW dermatomyositis; scleroderma; asthma; psoriasis; eczematous dermatitis;  
 KW Kaposi's sarcoma; inflammatory bowel disease; proliferative disorder.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT MISC-difference 10 /note="any amino acid"  
 FT  
 XX  
 PN MO200015252-A1.  
 XX  
 PD 23-MAR-2000.  
 XX  
 PF 09-SEP-1999; 99MO-US18842.  
 XX  
 PR 14-SEP-1998; 98US-0100266.  
 XX  
 PA (UYMA-) UNIV MARYLAND BALTIMORE.  
 XX  
 PI Fasano A, Szelein MB, Lu R, Tanner MK;  
 XX  
 DR WPI; 2000-271257/23.  
 XX  
 PT Suppression of antigen presenting cell mediated lymphocyte  
 PT proliferation, by administering a Zot-related immunoregulator useful  
 PT for treating immune-related disorders, immune system rejection  
 PT subsequent to tissue or organ transplantation  
 XX  
 PS Disclosure; Page 26; 95pp; English.  
 XX



CC human chimeric antibody chMint5 was deduced from a cDNA clone  
 CC (T33445) obt'd. by PCR amplification of murine MInt5 hybridoma DSM  
 CC ACC2150 cDNA. chMint5 comprises MInt5 VH and VL regions fused  
 CC to human C gamma1 and C kappa regions. Constructs were expressed in  
 CC CHO cell transfectants. chMint5 shows lower immunogenicity than  
 CC MInt5 when administered to humans. It can be used in diagnostic  
 CC assays or used to produce immunotoxins or immunocytokines useful  
 CC for tumour therapy.

XX Sequence 120 AA;

Query Match 88.1%; Score 37; DB 17; Length 120;

Best Local Similarity 87.5%; Pred. No. 9.7;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGLVQPG 8  
 || |||||

Db 8 ggalvpg 15

RESULT 14

B53648 B53648 standard; Protein; 121 AA.

XX B53648;

DT 09-MAR-2001 (first entry)

DE Human colon cancer antigen protein sequence SEQ ID NO:1188.

XX Human; colon cancer; colon cancer antigen; diagnosis; detection;  
 KW identification; cytotoxic; cardiotoxic; neuroprotective; vulnery;  
 KW immunomodulatory; muscular; gynecological; gastrointestinal;  
 KW nephrotropic; antiinfective; antibacterial; gene therapy; wound;  
 KW neural disorder; immune system disorder; muscular disorder;  
 KW reproductive disorder; gastrointestinal disorder; renal disorder;  
 KW infectious disease; cardiovascular disorder.

XX Homo sapiens.

XX MO20005351-A1.

XX 21-SEP-2000.

XX 08-MAR-2000; 2000WO-US05883.

XX 12-MAR-1999; 99US-0124270.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Rosen CA, Ruben SM;

XX WPI: 2000-587534/55.

XX N-PSDB; C98405.

XX Colon cancer associated gene sequences, referred to as colon cancer  
 PT antigens, useful for the treatment, prevention, and diagnosis of colon  
 PT disorders such as colon cancer -

XX Claim 11; Page 1770; 2104pp; English.

XX C97991 to C98763 encode the human colon cancer associated proteins,  
 CC called human colon cancer antigens, given in B53234 to B54006. The human  
 CC colon cancer antigens can have cytosolic, cardiolitic, muscular;  
 CC neuroprotective, immunomodulatory, gynecological, gastrointestinal,  
 CC vulnery, nephrotropic, antiinfective and antibacterial activities, and  
 CC can be used in gene therapy. The colon cancer antigen polynucleotides,  
 CC proteins and antibodies to the proteins are useful for the prevention,  
 CC treatment and diagnosis of colon disorders, such as colon cancer. The  
 CC polynucleotides may be used in diagnostics and research, such as for  
 CC chromosome identification, and as hybridisation probes. The proteins  
 CC may also be used to prevent diseases such as neural disorders, immune

CC system disorders, muscular disorders, reproductive disorders,  
 CC gastrointestinal disorders, wounds, renal disorders, infectious  
 CC diseases, and cardiovascular disorders. C98764 to C98772 and B54007  
 CC represent sequences used in the exemplification of the present  
 CC invention.

XX Sequence 121 AA;

Query Match 88.1%; Score 37; DB 21; Length 121;

Best Local Similarity 87.5%; Pred. No. 9.8;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGLVQPG 8  
 || |||||

Db 11 ggalvpg 18

RESULT 15

R54811 R54811 standard; Protein; 123 AA.

XX R54811;

DT 18-OCT-1994 (first entry)

DE SPA-reactive VH region SpA2-08.

XX SPA domain D; Ig binding region; B-cell superantigen; SAg;

KW superantigen; heavy chain variable region; VH3 restricted antibody;  
 KW VH; protein-A; SpA2-08; combinatorial library; B-lymphocyte;  
 KW vaccine.

XX Homo sapiens.

XX WO9409818-A.

XX 11-MAY-1994.

XX 29-OCT-1993; 93WO-US10555.

XX 30-OCT-1992; 92US-0969936.

XX (REGC ) UNIV CALIFORNIA.

XX Silverman GJ;

XX WPI: 1994-167127/20.

XX N-PSDB; Q64851.

XX Stimulating prodn. of variable region gene family restricted  
 PT antibodies - through B-cell super-antigen vaccination  
 PS Disclosure; Page 103; 130pp; English.

XX A B-cell superantigen (sAg) is a fragment of SpA D domain that  
 CC specifically binds the Fab portion of variable region restricted  
 CC antibodies. The sAg is used to enhance production of VH, especially  
 CC VH3, restricted Abs. During attempts to identify sAgs, aa sequences  
 CC (R54784-801) of H chains from Ig reactive with mod-SpA, and aa and  
 CC DNA sequences (R54802-16, Q64842-56) of VH regions of SpA binders  
 CC obtained from combinatorial libraries were determined.

XX Sequence 123 AA;

Query Match 88.1%; Score 37; DB 15; Length 123;

Best Local Similarity 87.5%; Pred. No. 10;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGLVQPG 8  
 || |||||

Db 8 ggalvpg 15

Wed Jun 13 15:00:02 2001

pct-us01-05825a-19.rag

Page 9

Search completed: June 13, 2001, 14:14:39  
Job time: 380 sec

---

*This Page Blank (uspto)*



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

## OM protein - protein search, using sw model

Run on: June 13, 2001, 14:10:47 ; Search time 130.61 Seconds  
(Without alignments)  
4.209 Million cell updates/sec

Title: PCT-US01-05825a-19

Perfect score: 42

Sequence: 1 GGLVQPG 8

## Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 198801 seqs, 68722935 residues

Total number of hits satisfying chosen parameters: 198801

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

## Database :

1: PIR\_67:\*  
2: PIR1:\*  
3: PIR3:\*  
4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	42	100.0	427	2 D83347	probable aminotran
2	37	88.1	116	1 AIH02P	Ig heavy chain V-I
3	36	85.7	98	2 S26929	Ig heavy chain V-I
4	36	85.7	110	2 S36282	Ig heavy chain V-I
5	36	85.7	120	2 S44111	Ig heavy chain V-D
6	36	85.7	324	2 S61447	cellulase (EC 3.2.
7	36	85.7	489	2 T06350	cellulase (EC 3.2.
8	35	83.3	123	2 PT0380	Ig heavy chain V-I
9	34	81.0	32	2 F31485	Ig heavy chain V-I
10	34	81.0	36	2 D31485	Ig heavy chain V-I
11	34	81.0	38	2 S33402	Ig heavy chain V-I
12	34	81.0	59	2 A27606	Ig heavy chain V-a
13	34	81.0	59	2 S36381	Ig heavy chain V-I
14	34	81.0	60	2 S36382	Ig heavy chain V-I
15	34	81.0	82	2 C34964	Ig heavy chain pre
16	34	81.0	94	2 D25913	Ig heavy chain V-I
17	34	81.0	94	2 I67528	CD33 antigen homol
18	34	81.0	95	2 I67527	CD33 antigen homol
19	34	81.0	97	2 PI0872	Ig heavy chain V-I
20	34	81.0	97	2 S26885	Ig heavy chain V-I
21	34	81.0	97	2 S26886	Ig heavy chain V-I
22	34	81.0	97	2 S26890	Ig heavy chain V-I
23	34	81.0	97	2 S46462	Ig heavy chain V-I
24	34	81.0	97	2 S26895	Ig heavy chain V-I
25	34	81.0	97	2 S24855	Ig heavy chain V-I
26	34	81.0	98	2 PL0121	Ig heavy chain V-I
27	34	81.0	98	2 PH0874	Ig heavy chain V-I
28	34	81.0	98	2 PL0123	Ig heavy chain V-I
29	34	81.0	98	2 S26896	Ig heavy chain V-I

30	34	81.0	98	2 S29545	Ig heavy chain V-I
31	34	81.0	98	2 S26927	Ig heavy chain V-I
32	34	81.0	98	2 S26932	Ig heavy chain V-I
33	34	81.0	98	2 S26891	Ig heavy chain V-I
34	34	81.0	98	2 S26894	Ig heavy chain V-I
35	34	81.0	98	2 S26889	Ig heavy chain V-I
36	34	81.0	98	2 S26933	Ig heavy chain V-I
37	34	81.0	98	2 S26934	Ig heavy chain V-I
38	34	81.0	100	2 PL0122	Ig heavy chain V-I
39	34	81.0	100	2 S69896	Ig heavy chain V-I
40	34	81.0	100	2 S26925	Ig heavy chain V-I
41	34	81.0	100	2 S26926	Ig heavy chain V-I
42	34	81.0	101	2 PU0003	Ig heavy chain V-I
43	34	81.0	101	2 PU0002	Ig heavy chain V-I
44	34	81.0	105	2 S38488	Ig heavy chain V-I
45	34	81.0	105	2 PL0255	Ig heavy chain V-I

## ALIGNMENTS

RESULT 1  
D83347  
Probable aminotransferase PA2394 [Imported] - Pseudomonas aeruginosa (strain PA01)  
C:Species: Pseudomonas aeruginosa  
C>Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
C:Accession: D83347  
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Muzoguchi, S.D.; Warren, P.; Hickey, M.J.;  
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lardig, K.; L  
.; Lory, S.; Olson, M.V.  
Nature 406, 959-964, 2000  
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa  
A:Reference number: A62950; MUID:20437337  
A:Accession: D83347  
A:Status: Preliminary  
A:Molecule type: DNA  
A:Residues: 1-427 <STO>  
A:Cross-references: GB:AE004665; GB:AE004091; NID:99948426; PIDN:AA05782.1; GSPDB:GN  
A:Experimental source: strain PA01  
C:Genetics:  
A:Gene: PA2394

Query Match 100.0%; Score 42; DB 2; Length 427;  
Best Local Similarity 100.0%; Pred. No. 0.98;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGLVQPG 8  
DB 131 GGLVQPG 138

RESULT 2  
AIH02P  
Ig heavy chain V-III region (zap) - human (tentative sequence)  
C:Species: Homo sapiens (man)  
C>Date: 24-Apr-1984 #sequence\_revision 24-Apr-1984 #text\_change 31-Mar-2000  
C:Accession: A02061  
R:Capra, J.D.; Kehoe, J.M.  
Proc. Natl. Acad. Sci. U.S.A. 71, 845-848, 1974  
A:Title: Variable region sequences of five human immunoglobulin heavy chains of the V  
A:Reference number: A93794; MUID:74142702  
A:Accession: A02061  
A:Molecule type: Protein  
A:Residues: 1-116 <CAP>  
C:Comment: This chain was isolated from an IgM myeloma protein.  
C:Genetics:  
A:Gene: GDB:IGHV  
A:Cross-references: GDB:128528; OMIM:147070  
A:Map position: 14q32.33-14q32.33  
C:Superfamily: Immunoglobulin V region; Immunoglobulin homology  
C:Keywords: heterotrimer; immunoglobulin  
F15-98/Domain: Immunoglobulin homology <IMM>

F:22-96/Disulfide bonds: #status predicted

Query Match 88.1%; Score 37; DB 1; Length 116;  
Best Local Similarity 87.5%; Pred. No. 2.7;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GGLVOPG 8  
|||  
Db 8 GGLVOPG 15

RESULT 3  
Ig heavy chain V region (DP-33) - human (fragment)  
S26929  
C:Species: Homo sapiens (man)  
C>Date: 22-Nov-1993 #sequence\_revision 10-Nov-1995 #text\_change 23-Jul-1999  
C:Accession: S26929

R:Tomlinson, I.M.; Walter, G.; Marks, J.D.; Llewellyn, M.B.; Winter, G.  
J. Mol. Biol. 227, 776-798, 1992  
A:Title: The repertoire of human germline V(H) sequences reveals about fifty groups of

A:Reference number: S26885; MUID:93021117

A:Accession: S26929  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-98 <TOM>  
A:Cross-references: EMBL:Z12335; NID:932889; PIDN:CAA78205.1; PID:932890  
C:Superfamily: immunoglobulin V region; immunoglobulin homology  
C:Keywords: heterotrimer; immunoglobulin  
F:15-98/Domain: immunoglobulin homology <IMM>

Query Match 85.7%; Score 36; DB 2; Length 98;  
Best Local Similarity 75.0%; Pred. No. 3.6;  
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 GGLVOPG 8  
|||  
Db 8 GGLVOPG 15

RESULT 4  
S36282  
Ig heavy chain V region (clone alpha-FOG1-H6) - human (fragment)

C:Species: Homo sapiens (man)  
C>Date: 03-Feb-1994 #sequence\_revision 03-Feb-1994 #text\_change 23-Jul-1999  
C:Accession: S36282

R:Griffiths, A.D.; Malmqvist, M.; Marks, J.D.; Bye, J.M.; Embleton, M.J.; McCafferty, J.  
EMBO J. 12, 725-734, 1993

A:Title: Human anti-self antibodies with high specificity from phage display libraries.  
A:Reference number: S36256; MUID:93178448

A:Accession: S36282  
A:Status: preliminary; nucleic acid sequence not shown

A:Molecule type: mRNA

A:Cross-references: EMBL:Z18824; NID:933111; PIDN:CAA79276.1; PID:9339891  
C:Superfamily: immunoglobulin V region; immunoglobulin homology

C:Keywords: heterotrimer; immunoglobulin  
F:15-98/Domain: immunoglobulin homology <IMM>

Query Match 85.7%; Score 36; DB 2; Length 110;  
Best Local Similarity 87.5%; Pred. No. 4.1;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GGLVOPG 8  
|||  
Db 8 GGLVOPG 15

RESULT 5  
S4411  
Ig heavy chain V-D-J region - human

C:Species: Homo sapiens (man)

C>Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 23-Jul-1999

C:Accession: S4411  
R:Hawkins, R.E.; Zhu, D.; Ovecka, M.; Winter, G.; Hamblin, T.J.; Stevenson, F.K.

Submitted to the EMBL Data Library, March 1994  
A:Description: Idiotypic vaccination against human B-cell lymphoma: rescue of variabl

A:Reference number: S44105

A:Accession: S4411  
A:Status: preliminary

A:Molecule type: DNA  
A:Residues: 1-120 <HAM>

A:Cross-references: EMBL:Z31387; NID:9472965; PIDN:CAA83262.1; PID:940522  
C:Superfamily: immunoglobulin V region; immunoglobulin homology

C:Keywords: heterotrimer; immunoglobulin  
F:15-98/Domain: immunoglobulin homology <IMM>

Query Match 85.7%; Score 36; DB 2; Length 120;  
Best Local Similarity 75.0%; Pred. No. 4.4;  
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 GGLVOPG 8  
|||  
Db 8 GGLVOPG 15

RESULT 6  
S61447  
cellulase (EC 3.2.1.4) CX3 - pepper (fragment)

C:Species: Capsicum annuum (pepper)  
C>Date: 24-Jul-1998 #sequence\_revision 24-Jul-1998 #text\_change 08-Dec-2000  
C:Accession: S61447

R:Ferrarese, L.; Trinolli, L.; Moretto, P.; Polverino de Laureto, P.; Raschio, N.; Ca  
Plant Mol. Biol. 29, 735-747, 1995

A:Title: Differential ethylene-inducible expression of cellulase in pepper plants.  
A:Reference number: S61445; MUID:96128016

A:Accession: S61447  
A:Status: nucleic acid sequence not shown

A:Molecule type: mRNA

A:Residues: 1-324 <FER>  
A:Cross-references: EMBL:X83711

C:Superfamily: Arabidopsis membrane-anchored cellulase KOR  
C:Keywords: glycosidase; hydrolase

Query Match 85.7%; Score 36; DB 2; Length 324;  
Best Local Similarity 75.0%; Pred. No. 12;  
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 GGLVOPG 8  
|||  
Db 255 GGLVOPG 262

RESULT 7  
T06350  
cellulase (EC 3.2.1.4) Cel2 precursor - tomato

N:Alternate names: endo-1,4-beta-glucanase  
C:Species: Lycopersicon esculentum (tomato)  
C>Date: 30-Apr-1999 #sequence\_revision 30-Apr-1999 #text\_change 08-Dec-2000

R:Ashbrook, C.C.; Gonzalez-Bosch, C.; Bennett, A.B.  
Plant Cell 6, 1485-1493, 1994

A:Title: Two divergent endo-beta-1,4-glucanase genes exhibit overlapping expression 1  
A:Reference number: Z15614; MUID:95086382

A:Accession: T06350  
A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA  
A:Residues: 1-489 <LAS>

A:Cross-references: EMBL:U03055; NID:9531904; PIDN:AAA69909.1; PID:9531905  
A:Experimental source: strain Castlemar; tissue-type pericarp

C:Genetics:  
A:Gene: Cel2

C:Function:  
A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as ce  
A:Pathway: cellulose degradation  
C:Superfamily: Arabidopsis membrane-anchored cellulase KOR  
C:Keywords: glycosidase; hydrolase; polysaccharide degradation  
F:1.23/Domain: signal sequence #status predicted <SIG>  
F:2.4/89/Product: cellulase 2 #status predicted <MAT>

Query Match 85.7%: Score 36; DB 2; Length 489;  
Best Local Similarity 75.0%: Pred. No. 18;  
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 GGLVOPG 8  
|||:||||  
Db 327 GGLVOPG 334

## RESULT 8

Ig heavy chain V region (S107/VH11 group 2-18) - mouse (fragment)  
PT0380  
C:Species: Mus musculus (house mouse)  
C:Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 07-May-1999  
C:Accession: PT0380  
R:Behar, S.M.; Lustgarten, D.L.; Corbet, S.; Scharff, W.D.  
J. Exp. Med. 173, 731-741, 1991  
A:Title: Characterization of somatically mutated S107 VH11-encoded anti-DNA autoantibodi  
A:Reference number: PT0376; MUID:9147903  
A:Accession: PT0380  
A:Molecule type: DNA  
A:Residues: 1-123 <BEH>  
C:Superfamily: Immunoglobulin V region; Immunoglobulin homology  
C:Keywords: heterotetramer; Immunoglobulin  
F:15-100/Domain: Immunoglobulin homology <IMM>

Query Match 83.3%: Score 35; DB 2; Length 123;  
Best Local Similarity 87.5%: Pred. No. 7.2;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GGLVOPG 8  
|||:||||  
Db 8 GGNLVOPG 15

## RESULT 9

Ig heavy chain V region (10-25) - mouse (fragment)  
F31485  
C:Species: Mus musculus (house mouse)  
C:Date: 31-Jul-1989 #sequence\_revision 31-Jul-1989 #text\_change 30-May-1997  
C:Accession: F31485  
R:Bedzyk, W.D.; Johnson, L.S.; Riordan, G.S.; Voss Jr., E.W.  
J. Biol. Chem. 264, 1565-1569, 1989  
A:Title: Comparison of variable region primary structures within an anti-fluorescein 1d  
A:Reference number: A31485; MUID:89109167  
A:Accession: F31485  
A>Status: preliminary  
A:Molecule type: protein  
A:Residues: 1-32 <BED>  
C:Superfamily: Immunoglobulin V region; Immunoglobulin homology  
C:Keywords: heterotetramer; Immunoglobulin

Query Match 81.0%: Score 34; DB 2; Length 32;  
Best Local Similarity 87.5%: Pred. No. 3;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GGLVOPG 8  
|||:||||  
Db 8 GGLVOPG 15

## RESULT 10

## D31485

Ig heavy chain V regions (3-24 and 5-14) - mouse (fragment)  
C:Species: Mus musculus (house mouse)  
C:Date: 31-Jul-1989 #sequence\_revision 31-Jul-1989 #text\_change 30-May-1997  
C:Accession: D31485  
R:Bedzyk, W.D.; Johnson, L.S.; Riordan, G.S.; Voss Jr., E.W.  
J. Biol. Chem. 264, 1565-1569, 1989  
A:Title: Comparison of variable region primary structures within an anti-fluorescein  
A:Reference number: A31485; MUID:89109167  
A:Accession: D31485  
A>Status: preliminary  
A:Molecule type: protein  
A:Residues: 1-36 <BED>  
C:Superfamily: Immunoglobulin V region; Immunoglobulin homology  
C:Keywords: heterotetramer; Immunoglobulin

Query Match 81.0%: Score 34; DB 2; Length 36;  
Best Local Similarity 87.5%: Pred. No. 3.3;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GGLVOPG 8  
|||:||||  
Db 8 GGLVOPG 15

## RESULT 11

Ig heavy chain V region - mouse (fragment)  
S33402  
C:Species: Mus musculus (house mouse)  
C:Date: 19-Mar-1997 #sequence\_revision 24-Jul-1997 #text\_change 17-Jul-1998  
C:Accession: S33402; S36385  
R:Ketleborough, C.A.; Saldanha, J.; Ansell, K.H.; Bendig, M.M.  
Eur. J. Immunol. 23, 206-211, 1993  
A:Title: Optimization of primers for cloning libraries of mouse immunoglobulin genes  
A:Reference number: S33391; MUID:93122092  
A:Accession: S33402  
A:Molecule type: mRNA  
A:Residues: 1-38 <KET>  
A:Cross-references: EMBL:X73019  
A:Experimental source: strain BALB/c  
R:Ansell, K.H.  
submitted to the EMBL Data Library, April 1993  
A:Reference number: S36376  
A:Accession: S36385  
A:Molecule type: mRNA  
A:Residues: 1-30 <ANS>  
A:Cross-references: EMBL:X73019  
C:Superfamily: Immunoglobulin V region; Immunoglobulin homology  
C:Keywords: heterotetramer; Immunoglobulin  
F:1-38/Domain: Immunoglobulin homology (fragment) <IMM>

Query Match 81.0%: Score 34; DB 2; Length 38;  
Best Local Similarity 87.5%: Pred. No. 3.5;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GGLVOPG 8  
|||:||||  
Db 6 GGLVOPG 13

## RESULT 12

Ig heavy chain V-a region (p26.9p2) - rabbit (fragments)  
A27606  
C:Species: Oryctolagus cuniculus (domestic rabbit)  
C:Date: 08-Mar-1989 #sequence\_revision 30-Jun-1991 #text\_change 16-Aug-1996  
C:Accession: A27606  
R:Currier, S.J.; Gallarda, J.L.; Knight, K.L.  
J. Immunol. 140, 1651-1659, 1988  
A:Title: Partial molecular genetic map of the rabbit V-H chromosomal region.  
A:Reference number: A27606; MUID:88154464  
A:Accession: A27606

A:Molecule type: DNA  
A:Residues: 1-59 <CUR>  
A:Cross-references: GB:M19706  
A>Note: the authors translated the codon TCC for residue 28 as Asp  
C:Genetics:  
A:introns: 30/3  
C:Superfamily: Immunoglobulin V region; immunoglobulin homology  
C:Keywords: heterotetramer; immunoglobulin

Query Match 81.0%; Score 34; DB 2; Length 59;  
Best Local Similarity 87.5%; Pred. No. 5.5;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GGLVOPG 8  
DB 8 GGGLVOPG 15

## RESULT 13

S36381  
Ig heavy chain V region - mouse (fragment)  
C:Species: Mus musculus (house mouse)  
C>Date: 19-Mar-1997 #sequence\_revision 29-Aug-1997 #text\_change 23-Jul-1999  
C:Accession: S36381; S33396  
R:Ansell, K.H.  
submitted to the EMBL Data Library, April 1993  
A:Reference number: S36376  
A:Accession: S36381  
A:Molecule type: mRNA  
A:Residues: 1-59 <ANS>  
A:Cross-references: EMBL:X73012; NID:g295879; PIDN:CA51498.1; PID:g939941  
R:Kettelborough, C.A.; Saldanha, J.; Ansell, K.H.; Bendig, M.M.  
Eur. J. Immunol. 23, 206-211, 1993  
A:Title: Optimization of primers for cloning libraries of mouse immunoglobulin genes  
A:Reference number: S33391; MUID:93122092  
A:Accession: S33396  
A:Molecule type: mRNA  
A:Residues: 3-43 <KET>  
A:Cross-references: EMBL:X73012  
C:Superficial source: Strain BALB/c  
C:Superfamily: Immunoglobulin V region; immunoglobulin homology  
C:Keywords: heterotetramer; immunoglobulin  
F:14-59/Domain: immunoglobulin homology (fragment) <IMM>

Query Match 81.0%; Score 34; DB 2; Length 59;  
Best Local Similarity 87.5%; Pred. No. 5.5;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GGLVOPG 8  
DB 10 GGGLVOPG 17

## RESULT 14

S36382  
Ig heavy chain V region - mouse (fragment)  
C:Species: Mus musculus (house mouse)  
C>Date: 19-Mar-1997 #sequence\_revision 29-Aug-1997 #text\_change 23-Jul-1999  
C:Accession: S36382; S33397  
R:Ansell, K.H.  
submitted to the EMBL Data Library, April 1993  
A:Reference number: S36376  
A:Accession: S36382  
A:Molecule type: mRNA  
A:Residues: 1-60 <ANS>  
A:Cross-references: EMBL:X73011; NID:g295878; PIDN:CA51497.1; PID:g939940  
R:Kettelborough, C.A.; Saldanha, J.; Ansell, K.H.; Bendig, M.M.  
Eur. J. Immunol. 23, 206-211, 1993  
A:Title: Optimization of primers for cloning libraries of mouse immunoglobulin genes  
A:Reference number: S33391; MUID:93122092  
A:Accession: S33397

A:Molecule type: mRNA  
A:Residues: 1-33 <KET>  
A:Cross-references: EMBL:X73011  
A:Experimental source: Strain BALB/c  
C:Superfamily: Immunoglobulin V region; immunoglobulin homology  
C:Keywords: heterotetramer; immunoglobulin  
F:8-60/Domain: immunoglobulin homology (fragment) <IMM>

Query Match 81.0%; Score 34; DB 2; Length 60;  
Best Local Similarity 87.5%; Pred. No. 5.6;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GGLVOPG 8  
DB 1 GGGLVOPG 8

## RESULT 15

C34964  
Ig heavy chain precursor V-II region (Ab21) - human  
C:Species: Homo sapiens (man)  
C>Date: 07-Sep-1990 #sequence\_revision 07-Sep-1990 #text\_change 09-May-1997  
C:Accession: C34964  
R:Sanz, I.; Casali, P.; Thomas, J.W.; Nockins, A.L.; Capra, J.D.  
J. Immunol. 142, 4054-4061, 1989  
A:Title: Nucleotide sequences of eight human natural autoantibody V-H regions reveals  
A:Reference number: A92830; MUID:89235232  
A:Accession: C34964  
A:Molecule type: mRNA  
A:Residues: 1-82 <SAN>  
A:Cross-references: GB:M26996  
A>Note: the sequence shown differs from the authors' translation after residue 56  
C:Superfamily: immunoglobulin V region; immunoglobulin homology  
C:Keywords: heterotetramer; immunoglobulin

Query Match 81.0%; Score 34; DB 2; Length 82;  
Best Local Similarity 87.5%; Pred. No. 7.6;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GGLVOPG 8  
DB 27 GGGLVOPG 34

Search completed: June 13, 2001, 14:10:47  
Job time: 148 sec



*This Page Blank (uspto)*

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 13, 2001, 14:21:48 ; Search time 74.44 Seconds

(without alignments)  
3.681 Million cell updates/sec

Title: PCT-US01-05825A-19  
Perfect score: 42  
Sequence: 1 GGLVQPG 8

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 3425486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_39:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	37	88.1	116	1 HV30_HUMAN	P01778 homo sapien
2	34	81.0	111	1 HV35_MOUSE	P01804 mus musculu
3	34	81.0	113	1 HV27_MOUSE	P01796 mus musculu
4	34	81.0	113	1 HV28_MOUSE	P01797 mus musculu
5	34	81.0	113	1 HV29_MOUSE	P01798 mus musculu
6	34	81.0	113	1 HV30_MOUSE	P01799 mus musculu
7	34	81.0	113	1 HV31_MOUSE	P01800 mus musculu
8	34	81.0	113	1 HV34_MOUSE	P01803 mus musculu
9	34	81.0	115	1 HV32_MOUSE	P01801 mus musculu
10	34	81.0	115	1 HV33_MOUSE	P01802 mus musculu
11	34	81.0	115	1 HV3D_HUMAN	P01765 homo sapien
12	34	81.0	116	1 HV36_MOUSE	P01806 mus musculu
13	34	81.0	116	1 HV3R_HUMAN	P01779 homo sapien
14	34	81.0	116	1 HV3T_HUMAN	P01781 homo sapien
15	34	81.0	117	1 HV17_MOUSE	P01786 mus musculu
16	34	81.0	117	1 HV3C_HUMAN	P01764 homo sapien
17	34	81.0	117	1 HV3O_HUMAN	P01776 homo sapien
18	34	81.0	117	1 HV41_MOUSE	P01811 mus musculu
19	34	81.0	117	1 HV54_MOUSE	P18525 mus musculu
20	34	81.0	118	1 HV39_MOUSE	P01809 mus musculu
21	34	81.0	118	1 HV3V_HUMAN	P80419 homo sapien
22	34	81.0	119	1 HV37_MOUSE	P01807 mus musculu
23	34	81.0	119	1 HV38_MOUSE	P01808 mus musculu
24	34	81.0	119	1 HV3M_HUMAN	P01774 homo sapien
25	34	81.0	119	1 HV3N_HUMAN	P01775 homo sapien
26	34	81.0	119	1 HV3P_HUMAN	P01777 homo sapien
27	34	81.0	119	1 HV4O_MOUSE	P01810 mus musculu
28	34	81.0	120	1 HV3E_HUMAN	P01782 homo sapien
29	34	81.0	120	1 HV3U_HUMAN	P01786 homo sapien
30	34	81.0	122	1 HV2O_MOUSE	P01789 mus musculu
31	34	81.0	122	1 HV21_MOUSE	P01790 mus musculu
32	34	81.0	123	1 HV18_MOUSE	P01787 mus musculu
33	34	81.0	123	1 HV19_MOUSE	P01788 mus musculu

34	34	81.0	123	1 HV22_MOUSE	P01791 mus musculu
35	34	81.0	123	1 HV23_MOUSE	P01792 mus musculu
36	34	81.0	123	1 HV24_MOUSE	P01793 mus musculu
37	34	81.0	123	1 HV25_MOUSE	P01794 mus musculu
38	34	81.0	136	1 HV16_MOUSE	P01783 mus musculu
39	34	81.0	142	1 HV01_RAT	P01805 rattus norv
40	34	81.0	144	1 HV26_MOUSE	P01795 mus musculu
41	34	81.0	238	1 Y457_CHITR	O84463 chlamydia t
42	34	81.0	403	1 YCOA_STNP7	P42460 synecchococ
43	34	81.0	1183	1 DRPL_RAT	P54258 rattus norv
44	34	81.0	1185	1 DRPL_HUMAN	P54259 homo sapien
45	33	78.6	115	1 HV3F_HUMAN	P01767 homo sapien

## ALIGNMENTS

```

RESULT 1
ID HV30_HUMAN          STANDARD;          PRT;          116 AA.
AC P01778;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE IG HEAVY CHAIN V-III REGION ZAP.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE.
RX MEDLINE=74142702; PubMed=4522793;
RA Capra J.D., Kehoe J.M.;
RT "Variable region sequences of five human immunoglobulin heavy chains
RT of the VH3 subgroup: definitive identification of four heavy chain
RT hypervariable regions."
RL Proc. Natl. Acad. Sci. U.S.A. 71:845-848(1974).
CC -I- MISCELLANEOUS: THIS CHAIN WAS ISOLATED FROM AN IGAL MYELOMA
CC PROTEIN.
DR PIR; A02061; ALH02P.
DR HSP; P01772; 21G2.
DR InterPro; IPR003006; -.
DR Pfam; PF00047; 1g; 1.
KW Immunoglobulin V region.
FT NON_TER 116
SQ
SEQUENCE 116 AA; 12582 MW; 892F8C217CEC9865 CRC64;

Query Match      88.1%; Score 37; DB 1; Length 116;
Best Local Similarity 87.5%; Pred. No. 14;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGLVQPG 8
   111111
DB 8 GGLVQPG 15

RESULT 2
ID HV35_MOUSE          STANDARD;          PRT;          111 AA.
AC P01804;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE IG HEAVY CHAIN V-III REGION HPC76 (FRAGMENT).
OS Mus musculus (Mouse)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=81013937; PubMed=6251474;
RA Bernard O., Gough N.M.;

```

RT "Nucleotide sequence of immunoglobulin heavy chain joining segments between translocated VH and mu constant regions genes.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 77:3630-3634(1980).  
 CC -1- MISCELLANEOUS: THE SEQUENCE OF THE FIRST 197 RESIDUES OF THE C REGION WAS ALSO DETERMINED & DIFFERS IN ONLY 3 POSITIONS FROM THE CORRESPONDING PORTION OF THE MOUSE MOPC 104E MU CHAIN.  
 CC PIR: A02074; MHMS76.  
 DR InterPro: IPR003006; -.  
 DR Pfam: PF00047; 1g.1.  
 DR Immunoglobulin V region.  
 FT NON\_TER 1 111  
 FT NON\_TER 111 111  
 SQ SEQUENCE 111 AA; 12304 MW; 0EDE98EC7348056A CRC64;

Query Match 81.0%; Score 34; DB 1; Length 111;  
 Best Local Similarity 87.5%; Pred. No. 5.3;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GGLVOPG 8  
 || |||||  
 Db 8 GGLVOPG 10

RESULT 3  
 HV27\_MOUSE STANDARD; PRT; 113 AA.  
 AC P01796;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE 1G HEAVY CHAIN V-III REGION A4.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE.  
 RX MEDLINE=78158406; PubMed=417344;  
 RA Vrana M., Rudikoff S., Potter M.;  
 RT "Sequence variation among heavy chains from Inulin-binding myeloma proteins.";  
 RT Proc. Natl. Acad. Sci. U.S.A. 75:1957-1961(1978).  
 CC -1- MISCELLANEOUS: THIS CHAIN WAS ISOLATED FROM A MYELOMA PROTEIN THAT BINDS INULIN.  
 CC PIR: A93818; AVMSAB.  
 DR InterPro: IPR003006; -.  
 DR Pfam: PF00047; 1g.1.  
 DR Immunoglobulin V region.  
 FT DISULFID 22 98  
 FT NON\_TER 113 113  
 FT NON\_TER 113 113  
 SQ SEQUENCE 113 AA; 12675 MW; 76658C121C598285 CRC64;

Query Match 81.0%; Score 34; DB 1; Length 113;  
 Best Local Similarity 87.5%; Pred. No. 5.3;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GGLVOPG 8  
 || |||||  
 Db 8 GGLVOPG 15

RESULT 4  
 HV28\_MOUSE STANDARD; PRT; 113 AA.  
 AC P01797;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE 1G HEAVY CHAIN V-III REGION U61.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE.  
 RX MEDLINE=78158406; PubMed=417344;  
 RA Vrana M., Rudikoff S., Potter M.;  
 RT "Sequence variation among heavy chains from Inulin-binding myeloma proteins.";  
 RT Proc. Natl. Acad. Sci. U.S.A. 75:1957-1961(1978).  
 CC -1- MISCELLANEOUS: THIS CHAIN WAS ISOLATED FROM A MYELOMA PROTEIN THAT BINDS INULIN.  
 CC PIR: B93818; AVMS61.  
 DR InterPro: IPR003006; -.  
 DR Pfam: PF00047; 1g.1.  
 DR Immunoglobulin V region.  
 FT DISULFID 22 98  
 FT NON\_TER 113 113  
 FT NON\_TER 113 113  
 SQ SEQUENCE 113 AA; 12671 MW; 76634C121C598285 CRC64;

Query Match 81.0%; Score 34; DB 1; Length 113;  
 Best Local Similarity 87.5%; Pred. No. 5.3;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GGLVOPG 8  
 || |||||  
 Db 8 GGLVOPG 15

RESULT 5  
 HV29\_MOUSE STANDARD; PRT; 113 AA.  
 AC P01798;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE 1G HEAVY CHAIN V-III REGION E109.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE.  
 RX MEDLINE=78158406; PubMed=417344;  
 RA Vrana M., Rudikoff S., Potter M.;  
 RT "Sequence variation among heavy chains from Inulin-binding myeloma proteins.";  
 RT Proc. Natl. Acad. Sci. U.S.A. 75:1957-1961(1978).  
 CC -1- MISCELLANEOUS: THIS CHAIN WAS ISOLATED FROM A MYELOMA PROTEIN THAT BINDS INULIN.  
 CC PIR: C93818; AVMS09.  
 DR InterPro: IPR003006; -.  
 DR Pfam: PF00047; 1g.1.  
 DR Immunoglobulin V region.  
 FT DISULFID 22 98  
 FT NON\_TER 113 113  
 FT NON\_TER 113 113  
 SQ SEQUENCE 113 AA; 12647 MW; EE50F2F20EDB129B CRC64;

Query Match 81.0%; Score 34; DB 1; Length 113;  
 Best Local Similarity 87.5%; Pred. No. 5.3;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GGLVOPG 8  
 || |||||  
 Db 8 GGLVOPG 15

RESULT 6  
 HV30\_MOUSE STANDARD; PRT; 113 AA.  
 ID HV30\_MOUSE  
 AC P01799;  
 DT 21-JUL-1986 (Rel. 01, Created)



DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DE 15-JUL-1999 (Rel. 38, Last annotation update)  
 ID IG HEAVY CHAIN V-III REGION ABE-47N.  
 OS Mus musculus (Mouse).  
 CC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
 DE NCBI\_Taxid=10090;  
 RN [1]  
 RP SEQUENCE.  
 RX MEDLINE=77134726; PubMed=402936;  
 RA Vrena M., Rudikoff S., Potter M.;  
 RT "Heavy-chain variable-region sequence from an inulin-binding myeloma  
 protein."  
 CC Blochemistry 16:1170-1175(1977).  
 DE -I- MISCELLANEOUS: THIS CHAIN WAS ISOLATED FROM A MYELOMA PROTEIN THAT  
 CC BINDS INULIN.  
 CC PIR: A90400; AVMSB7.  
 DR InterPro: IPR003006; -.  
 DR Pfam: PF00047; 1g; 1.  
 KW Immunoglobulin V region.  
 FT DISULFID 22 98  
 FT NON\_TER 113 113  
 SQ SEQUENCE 113 AA; 12675 MW; 76658C16C779845E CRC64;

Query Match 81.0%; Score 34; DB 1; Length 113;  
 Best Local Similarity 87.5%; Pred. No. 5.3;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GGLVQPG 8  
 |||||  
 DB 8 GGLVQPG 15

RESULT 7  
 ID HV31\_MOUSE STANDARD; PRT; 113 AA.  
 AC P01800;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 01-OCT-1993 (Rel. 27, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE IG HEAVY CHAIN V-III REGION T957.  
 OS Mus musculus (Mouse).  
 CC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
 DE NCBI\_Taxid=10090;  
 RN [1]  
 RP SEQUENCE.  
 RX MEDLINE=81216632; PubMed=6787122;  
 RA Rudikoff S., Potter M.;  
 RT "Immunoglobulin heavy chains from anti-inulin myeloma proteins:  
 RT evidence for a new heavy chain joining segment."  
 RL J. Immunol. 127:191-194(1981).  
 CC -I- MISCELLANEOUS: THIS CHAIN WAS ISOLATED FROM A MYELOMA PROTEIN THAT  
 CC BINDS INULIN.  
 CC PIR: A92810; AVMS57.  
 DR InterPro: IPR003006; -.  
 DR Pfam: PF00047; 1g; 1.  
 KW Immunoglobulin V region.  
 FT DISULFID 22 98  
 FT NON\_TER 113 113  
 SQ SEQUENCE 113 AA; 12732 MW; 26618F62B59859E CRC64;

Query Match 81.0%; Score 34; DB 1; Length 113;  
 Best Local Similarity 87.5%; Pred. No. 5.3;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GGLVQPG 8  
 |||||  
 DB 8 GGLVQPG 15

RESULT 8  
 ID HV34\_MOUSE STANDARD; PRT; 113 AA.  
 AC P01803;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE IG HEAVY CHAIN V REGION AMCL.  
 OS Mus musculus (Mouse).  
 CC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
 DE NCBI\_Taxid=10090;  
 RN [1]  
 RP SEQUENCE.  
 RX MEDLINE=81216632; PubMed=6787122;  
 RA Rudikoff S., Potter M.;  
 RT "Immunoglobulin heavy chains from anti-inulin myeloma proteins:  
 RT evidence for a new heavy chain joining segment."  
 RL J. Immunol. 127:191-194(1981).  
 CC -I- MISCELLANEOUS: THIS CHAIN WAS ISOLATED FROM A MYELOMA PROTEIN THAT  
 CC BINDS INULIN.  
 CC PIR: A02073; HVMSAM.  
 DR InterPro: IPR003006; -.  
 DR Pfam: PF00047; 1g; 1.  
 KW Immunoglobulin V region.  
 FT DISULFID 22 98  
 FT NON\_TER 113 113  
 SQ SEQUENCE 113 AA; 12691 MW; 7A6D906AA966E9E CRC64;

Query Match 81.0%; Score 34; DB 1; Length 113;  
 Best Local Similarity 87.5%; Pred. No. 5.3;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GGLVQPG 8  
 |||||  
 DB 8 GGLVQPG 15

RESULT 9  
 ID HV32\_MOUSE STANDARD; PRT; 115 AA.  
 AC P01801;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE IG HEAVY CHAIN V-III REGION J606.  
 OS Mus musculus (Mouse).  
 CC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
 DE NCBI\_Taxid=10090;  
 RN [1]  
 RP SEQUENCE.  
 RX MEDLINE=82099361; PubMed=6798111;  
 RA Johnson N., Stansard J., Paul L., Hood L.;  
 RT "The complete V domain amino acid sequences of two myeloma inulin-  
 RT binding proteins."  
 RL J. Immunol. 128:302-307(1982).  
 CC -I- MISCELLANEOUS: THIS CHAIN WAS ISOLATED FROM A MYELOMA PROTEIN THAT  
 CC BINDS INULIN.  
 CC PIR: G92811; AVMS06.  
 DR InterPro: IPR003006; -.  
 DR Pfam: PF00047; 1g; 1.  
 KW Immunoglobulin V region.  
 FT DISULFID 22 98  
 FT NON\_TER 115 115  
 SQ SEQUENCE 115 AA; 12810 MW; B67AD6638A121A5F CRC64;

Query Match 81.0%; Score 34; DB 1; Length 115;  
 Best Local Similarity 87.5%; Pred. No. 5.4;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GGLVOPG 8  
111111  
Db 8 GGLVOPG 15

## RESULT 10

HV33\_MOUSE STANDARD; PRT; 115 AA.  
ID HV33\_MOUSE  
AC P01802;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 21-JUL-1986 (Rel. 01, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE IG HEAVY CHAIN V-III REGION W3082.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE.  
RX MEDLINE=82099361; PubMed=679811;  
RA Johnson N., Stankard J., Paul L., Hood L.;  
RT "The complete V domain amino acid sequences of two myeloma Inulin-  
binding proteins."  
RT J. Immunol. 128:302-307(1982).  
CC -1- MISCELLANEOUS: THIS CHAIN WAS ISOLATED FROM A MYELOMA PROTEIN THAT  
BINDS INULIN.  
CC PIR: D92811; AVMS82.  
DR InterPro: IPR003006; -  
DR Pfam: PF00047; 19; 1.  
KW Immunoglobulin V region.  
FT DISULFID 22 98  
FT NON\_TER 115 115 BY SIMILARITY.  
SQ SEQUENCE 115 AA; 12887 MW; 9B4517648C121C5A CRC64;

Query Match 81.0%; Score 34; DB 1; Length 115;  
Best Local Similarity 87.5%; Pred. No. 5.4;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GGLVOPG 8  
111111  
Db 8 GGLVOPG 15

## RESULT 11

HV3D\_HUMAN STANDARD; PRT; 115 AA.  
ID HV3D\_HUMAN  
AC P01765;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 21-JUL-1986 (Rel. 01, Last sequence update).  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE IG HEAVY CHAIN V-III REGION TIL.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE.  
RX MEDLINE=78005528; PubMed=409716;  
RA Wang A.-C., Wang I.Y., Fudenberg H.H.;  
RT "Immunoglobulin structure and genetics. Identity between variable  
regions of a mu and a gamma2 chain."  
RT J. Biol. Chem. 252:7192-7199(1977).  
CC -1- MISCELLANEOUS: THE SEQUENCES OF THE V REGIONS OF THE HEAVY CHAINS  
OF IGM AND IGG2 ISOLATED FROM A SINGLE PATIENT WITH BICLONAL  
GAMMOPATHY ARE IDENTICAL. THEIR LIGHT CHAINS ARE APPARENTLY ALSO  
IDENTICAL.  
CC PIR: A02048; H3HUTL.  
DR InterPro: IPR003006; -  
DR Pfam: PF00047; 19; 1.  
KW Immunoglobulin V region.  
FT NON\_TER 115 115  
SQ SEQUENCE 115 AA; 12356 MW; 4DCC67D179F62326 CRC64;

Query Match 81.0%; Score 34; DB 1; Length 115;  
Best Local Similarity 87.5%; Pred. No. 5.4;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GGLVOPG 8  
111111  
Db 8 GGLVOPG 15

## RESULT 12

HV36\_MOUSE STANDARD; PRT; 116 AA.  
ID HV36\_MOUSE  
AC P01806;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 21-JUL-1986 (Rel. 01, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE IG HEAVY CHAIN V REGION 441 PRECURSOR.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=82059449; PubMed=6795591;  
RA Olio R., Aufray C., Sikorav J.-L., Rougeon F.;  
RT "Mouse heavy chain variable regions: nucleotide sequence of a  
germ-line VH gene segment."  
RT Nucleic Acids Res. 9:4099-4109(1981).  
CC -----

This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

CC EMBL: V00774; CAA24152.1; -  
CC PIR: A02076; HVMS44.  
CC HSSP: P01810; 2FBJ.  
DR InterPro: IPR003006; -  
DR Pfam: PF00047; 19; 1.  
KW Immunoglobulin V region; Signal.  
FT SIGNAL 1 18  
FT CHAIN 19 116 IG HEAVY CHAIN V REGION 441.  
FT NON\_TER 116 116  
SQ SEQUENCE 116 AA; 12911 MW; 6B4FF3EF9A95D9F3 CRC64;

Query Match 81.0%; Score 34; DB 1; Length 116;  
Best Local Similarity 87.5%; Pred. No. 5.5;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GGLVOPG 8  
111111  
Db 26 GGLVOPG 33

## RESULT 13

HV3R\_HUMAN STANDARD; PRT; 116 AA.  
ID HV3R\_HUMAN  
AC P01779;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 21-JUL-1986 (Rel. 01, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE IG HEAVY CHAIN V-III REGION TUR.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]

```

RP SEQUENCE.
RX MEDLINE=74142702; PubMed=4522793;
RA Capra J.D., Kehoe J.M.;
RT "Variable region sequences of five human immunoglobulin heavy chains
RT of the VH3 subgroup; definitive identification of four heavy chain
RT hypervariable regions."
RL Proc. Natl. Acad. Sci. U.S.A. 71:845-848(1974).
CC -1- MISCELLANEOUS: THIS CHAIN WAS ISOLATED FROM AN IGAI MYELOMA
CC PROTEIN.
DR PIR: A02062; A1HUTU.
DR InterPro: IPR003006; -.
DR Pfam: PF00047; 1g; 1.
DR Immunoglobulin V region.
FT NON_TER 116 116
SQ SEQUENCE 116 AA; 12431 MW; EB705F55A963F0C CRC64;

Query Match
Best Local Similarity 81.0%; Score 34; DB 1; Length 116;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGLVQPG 8
DB 8 GGLVQPG 15

RESULT 14
HV3T_HUMAN STANDARD; PRT; 116 AA.
ID HV3T_HUMAN
AC P01781;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE IG HEAVY CHAIN V-III REGION GAL.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE.
RX MEDLINE=75059123; PubMed=4803843;
RA Watanabe S., Barnikol H.U., Horn J., Hilschmann N.;
RT "The primary structure of a monoclonal Igm-immunoglobulin
RT (macroglobulin Gal.), II: the amino acid sequence of the H-chain (mu-
RT type), subgroup H III. Architecture of the complete Igm-molecule."
RL Hoppe-Seyler's Z. Physiol. Chem. 354:1505-1509(1973).
RN [2]
RP REVISION TO THE COMPOSITION OF 28-33.
RA Hilschmann N.;
RL Submitted (JUN-1975) to the PIR data bank.
CC -1- MISCELLANEOUS: THIS MU CHAIN WAS ISOLATED FROM A WALDENSTROM'S
CC MACROGLOBULIN.
DR PIR: A02064; M3HUGL.
DR HSSP: P01772; 2IG2.
DR InterPro: IPR003006; -.
DR Pfam: PF00047; 1g; 1.
DR Immunoglobulin V region.
FT NON_TER 116 116
SQ SEQUENCE 116 AA; 12730 MW; 2C67CA9AAAAA1282 CRC64;

Query Match
Best Local Similarity 81.0%; Score 34; DB 1; Length 116;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGLVQPG 8
DB 8 GGLVQPG 15

RESULT 15
HV17_MOUSE STANDARD; PRT; 117 AA.
ID HV17_MOUSE

```

```

AC P01786;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE IG HEAVY CHAIN V REGION MOPC 47A.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE.
RX MEDLINE=80049769; PubMed=115869;
RA Robinson E.A., Appella E.;
RT "Amino acid sequence of a mouse myeloma immunogloblin heavy chain
RT (MOPC 47 A) with a 100-residue deletion."
RL J. Biol. Chem. 254:11418-11430(1979).
CC -1- MISCELLANEOUS: THIS ALPHA CHAIN WAS ISOLATED FROM A MYELOMA
CC MOLECULES CONTAIN TWO LIGHT AND TWO HEAVY CHAINS AND LACK A
CC LINKED BY A DISULFIDE BOND. IN CONTRAST, NORMAL MOUSE IGA
CC PROTEIN THAT CONTAINS ONE LIGHT AND ONE HEAVY CHAIN PER MOLECULE,
CC MOLECULES CONTAIN TWO LIGHT AND TWO HEAVY CHAINS AND LACK A
CC LIGHT-HEAVY CHAIN DISULFIDE BOND.
DR PIR: A02069; A1MS47.
DR HSSP: P01789; 2MCP.
DR InterPro: IPR003006; -.
DR Pfam: PF00047; 1g; 1.
DR Immunoglobulin V region.
FT NON_TER 117 117
SQ SEQUENCE 117 AA; 12975 MW; 0C74BB8B154BDF4 CRC64;

Query Match
Best Local Similarity 81.0%; Score 34; DB 1; Length 117;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGLVQPG 8
DB 8 GGLVQPG 15

```

Search completed: June 13, 2001, 14:21:48  
Job time: 808 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 13, 2001, 14:20:36 ; Search time 225.85 seconds  
(without alignments)  
4.152 Million cell updates/sec

Title: PCT-US01-05825A-19  
Perfect score: 42  
Sequence: 1 GGLVQPG 8

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 374700 seqs, 117207915 residues  
Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: SPREMBL\_15:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_unclassified:\*  
13: sp\_vertebrate:\*  
14: sp\_virus:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	37	88.1	510	2	068278
2	36	85.7	325	10	043751
3	36	85.7	485	10	096547
4	36	85.7	489	10	042872
5	34	81.0	15	11	090V16
6	34	81.0	95	4	09ULB6
7	34	81.0	118	4	09UL91
8	34	81.0	118	4	09UL72
9	34	81.0	238	2	09PUT5
10	34	81.0	237	2	057998
11	34	81.0	272	1	09V2H0
12	34	81.0	323	13	073907
13	34	81.0	1175	11	035126
14	34	81.0	1175	11	P70200
15	34	81.0	1190	4	099621
16	33	78.6	186	10	09LHM3
17	33	78.6	222	1	09YDN2
18	33	78.6	238	2	09JQ12
19	33	78.6	251	2	09RSA9

20	33	78.6	263	2	09XAL6	09xal6 streptomyc
21	33	78.6	368	5	09YCE2	09yce2 aeropyrum p
22	33	78.6	396	5	020338	020338 caenorhabdi
23	33	78.6	444	5	010923	010923 caenorhabdi
24	33	78.6	494	10	09SVJ4	09svj4 arabidopsis
25	33	78.6	499	2	09SDP9	09sdp9 streptomyc
26	33	78.6	547	4	09S072	09s072 homo sapien
27	33	78.6	547	4	09NV09	09nv09 homo sapien
28	33	78.6	812	5	006452	006452 ephydalia m
29	33	78.6	904	5	09NEE8	09nee8 leishmania
30	33	78.6	1936	5	09U176	09u176 leishmania
31	32	76.2	167	5	006783	006783 haematobia
32	32	76.2	260	5	09VEH2	09veh2 drosophila
33	32	76.2	392	1	09YA73	09ya73 aeropyrum p
34	32	76.2	416	1	030223	030223 archaeoglob
35	32	76.2	456	1	057794	057794 pseudomonas
36	32	76.2	563	2	052169	052169 pseudomonas
37	32	76.2	563	2	006573	006573 pseudomonas
38	32	76.2	1050	4	09UF82	09uf82 homo sapien
39	32	76.2	1144	2	09JX22	09jx22 neisseria m
40	32	76.2	1144	2	09JX8	09jx8 neisseria m
41	32	76.2	1192	4	09Y2J3	09y2j3 homo sapien
42	32	76.2	1207	13	09PVF5	09pvf5 brachydanio
43	32	76.2	1248	4	095458	095458 homo sapien
44	32	76.2	1263	5	009971	009971 caenorhabdi
45	32	76.2	1557	2	086560	086560 streptomyc

## ALIGNMENTS

RESULT 1	
068278	PRELIMINARY; PRT; 510 AA.
AC	068278:
DT	01-AUG-1998 (TREMBLrel. 07, Created)
DT	01-AUG-1998 (TREMBLrel. 07, Last sequence update)
DT	01-NOV-1998 (TREMBLrel. 08, Last annotation update)
DE	ARF1 (EC 3.2.1.55) (ALPHA-L-ARABINOFURANOSIDASE) (ARABINOSIDASE).
GN	ARF1.
OS	Cytophaga xylanolytica.
OC	Bacteria; CFb group; Cytophagales; Cytophagaceae; Cytophaga.
OX	NCBI_TaxID=990;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=XM3;
RX	MEDLINE=96247330; PubMed=9572972;
RA	Kim K.S., Lilburn T.G., Renner M.J., Breznak J.A.;
RT	"Arf1 and arf1, two genes encoding alpha-L-arabinofuranosidases in
RL	Cytophaga xylanolytica.";
CC	Appl. Environ. Microbiol. 64:1919-1923(1998).
CC	-I- CATALYTIC ACTIVITY: HYDROLYSIS OF TERMINAL NON-REDUCING ALPHA-L-
DR	ARABINOFURANOSIDE RESIDUES IN ALPHA-L-ARABINOSIDES.
KW	EMBL; AF028018; AAC38456.1; -
SK	Hydrolase; Glycosidase.
SEQUENCE	510 AA; 57571 MW; BCE638E4DBC2A074 CRC64;
Query Match	88.1%; Score 37; DB 2; Length 510;
Best Local Similarity	75.0%; Pred. No. 33;
Matches	6; Conservative 2; Mismatches 0; Indels 0;
OY	1 GGLVQPG 8
DB	54 GGLVQPG 61
RESULT 2	
043751	PRELIMINARY; PRT; 325 AA.
ID	043751:
AC	043751:
DT	01-NOV-1996 (TREMBLrel. 01, Created)
DT	01-NOV-1996 (TREMBLrel. 01, Last sequence update)

DT 01-OCT-2000 (TREMBLrel. 15, last annotation update)  
DE CELLULOSE (EC 3.2.1.4) (ENDOGLUCANASE) (ENDO-1,4-BETA-GLUCANASE)  
OC (CARBOXYMETHYL CELLULOSE) (FRAGMENT).  
GN CCX3.  
OS Capsicum annuum (Bell pepper).  
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
OC Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I;  
OC Solanales; Solanaceae; Capsicum.  
OX NCBI\_TaxID=4072;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-CV. LONGUM; TISSUE-LEAF ABSCISSION ZONES;  
RX MEDLINE-96128016; PubMed-8541500;  
RA Ferrarese L., Trainotti L., Moretto P., Polymerio de lauro P.,  
RA Rasco N., Casadoro G.;  
RT "differential ethylene-inducible expression of cellulase in pepper  
plants.";  
RL Plant Mol. Biol. 29:735-747(1995).  
CC -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC  
LINKAGES IN CELLULOSE.  
DE EMBL; X83711; CAA58686.1; -.  
DR HSSP; P26221; 1TF4.  
DR MENDEL; 75759; Capan;1057;7579.  
DR INTERPRO; IPR001701; -.  
DR PFAM; PF00759; Glyco\_hydro\_9; 1.  
KW Hydrolyase; Glycosidase.  
FT NON\_TER 1  
FT NON\_TER 325  
SQ SEQUENCE 325 AA; 36087 MW; CE2040A6C7B17937 CRC64;

Query Match 85.7%; Score 36; DB 10; Length 325;  
Best Local Similarity 75.0%; Pred. No. 32;  
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 GGLVOPG 8  
DB 255 GGLVOPG 262

RESULT 3  
ID 096547 PRELIMINARY; PRT: 485 AA.  
AC 096547;  
DT 01-FEB-1997 (TREMBLrel. 02, Created)  
DT 01-FEB-1997 (TREMBLrel. 02, last sequence update)  
DE 01-OCT-2000 (TREMBLrel. 15, last annotation update)  
DE CELLULOSE (EC 3.2.1.4) (ENDOGLUCANASE) (ENDO-1,4-BETA-GLUCANASE)  
DE (CARBOXYMETHYL CELLULOSE).  
GN CCEU3.  
OS Capsicum annuum (Bell pepper).  
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
OC Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I;  
OC Solanales; Solanaceae; Capsicum.  
OX NCBI\_TaxID=4072;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-CV. LONGUM; TISSUE-ABSCISSION ZONE;  
RA Trainotti L., Ferrarese L., Casadoro G.;  
RT Submitted (Apr-1996) to the EMBL/GenBank/DBJ databases.  
CC -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC  
LINKAGES IN CELLULOSE.  
DE EMBL; X97189; CAA65827.1; -.  
DR HSSP; P26221; 1TF4.  
DR MENDEL; 7582; Capan;1057;7582.  
DR INTERPRO; IPR001701; -.  
DR PFAM; PF00759; Glyco\_hydro\_9; 1.  
DR PROSITE; PS00698; GLYCOSTYL\_HYDROL\_F9\_2; UNKNOWN\_1.  
KW Hydrolyase; Glycosidase.  
SQ SEQUENCE 485 AA; 53631 MW; ABBCB32E3FFA18A CRC64;

Query Match 85.7%; Score 36; DB 10; Length 485;

Best Local Similarity 75.0%; Pred. No. 49;  
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 GGLVOPG 8  
DB 324 GGLVOPG 331

RESULT 4  
ID 042872 PRELIMINARY; PRT: 489 AA.  
AC 042872;  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (TREMBLrel. 01, last sequence update)  
DE 01-OCT-2000 (TREMBLrel. 15, last annotation update)  
DE ENDO-1,4-BETA-GLUCANASE PRECURSOR (EC 3.2.1.4) (CELLULOSE)  
DE (ENDOGLUCANASE) (CARBOXYMETHYL CELLULOSE).  
GN CEL2.  
OS Lycopersicon esculentum (Tomato).  
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
OC Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I;  
OC Solanales; Solanaceae; Solanum.  
OX NCBI\_TaxID=4081;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-CASTLEMAKT; TISSUE-PERICARP;  
RX MEDLINE-95086382; PubMed-7994180;  
RA Lashbrook C.C., Gonzalez-Bosch C., Bennett A.B.;  
RT "two divergent endo-beta-1,4-glucanase genes exhibit overlapping  
expression in ripening fruit and abscising flowers.";  
RL Plant Cell 6:1485-1493(1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-CASTLEMAKT; TISSUE-PERICARP;  
RX Brumwell D.A., Lashbrook C.C., Bennett A.B.;  
RL Am. Chem. Soc. Symp. Ser. 566:100-129(1994).  
CC -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC  
LINKAGES IN CELLULOSE.  
DE EMBL; U13055; AAA69909.1; -.  
DR HSSP; P26221; 1TF4.  
DR MENDEL; 8802; Lycs;1057;8802.  
DR INTERPRO; IPR001701; -.  
DR PFAM; PF00759; Glyco\_hydro\_9; 1.  
DR PROSITE; PS00698; GLYCOSTYL\_HYDROL\_F9\_2; UNKNOWN\_1.  
KW Signal; Hydrolyase; Glycosidase.  
FT SIGNAL 1  
FT CHAIN 24  
FT CHAIN 489  
SQ SEQUENCE 489 AA; 54118 MW; FFADA8C4675F685F CRC64;

Query Match 85.7%; Score 36; DB 10; Length 489;  
Best Local Similarity 75.0%; Pred. No. 49;  
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 GGLVOPG 8  
DB 327 GGLVOPG 334

RESULT 5  
ID 090Y16 PRELIMINARY; PRT: 15 AA.  
AC 090Y16;  
DT 01-MAY-2000 (TREMBLrel. 13, Created)  
DT 01-MAY-2000 (TREMBLrel. 13, last sequence update)  
DE 01-OCT-2000 (TREMBLrel. 15, last annotation update)  
DE PROACTIN-BINDING PROTEIN (FRAGMENT).  
OS Rattus sp.  
OS Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10118;  
RN [1]  
RP SEQUENCE.

RX MEDLINE=95094032; PubMed=8000909;  
 RA Cohen H., Cohen O., Gagnon J.;  
 RT "Serum prolactin-binding protein (PRL-BP) of human and rat are  
 RT identified as IgG.";  
 RL C. R. Acad. Sci., III, Sci. vie 317:293-298(1994).  
 DR HSSP; P01789; IMCP.  
 SQ SEQUENCE 15 AA; 1469 MW; 35ED2512FF3FA369 CRC64;

Query Match 81.0%; Score 34; DB 11; Length 15;  
 Best Local Similarity 87.5%; Pred. No. 2.7;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GGLVOPG 8  
 || |||||  
 DB 8 GGLVOPG 15

## RESULT 6

O9ULB6 PRELIMINARY; PRT; 95 AA.  
 AC O9ULB6;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)  
 DE MYOSIN-REACTIVE IMMUNOGLOBULIN HEAVY CHAIN VARIABLE REGION (FRAGMENT).  
 GN VH.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA "Tange Y., Kayano H.;  
 RT "Human VH gene sequence.";  
 RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AB035268; BAA87067.1; -.  
 DR HSSP; P01772; 2F84.  
 DR INTERPRO; IPR003006; -.  
 DR PFAM; PF00047; 1g; 1.  
 FT NON\_TER 1  
 FT 95  
 FT NON\_TER 1  
 SQ SEQUENCE 95 AA; 10527 MW; 90A8CD1D22574A CRC64;

Query Match 81.0%; Score 34; DB 4; Length 95;  
 Best Local Similarity 87.5%; Pred. No. 20;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GGLVOPG 8  
 || |||||  
 DB 7 GGLVOPG 14

## RESULT 7

O9UL91 PRELIMINARY; PRT; 118 AA.  
 AC O9UL91;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)  
 DE MYOSIN-REACTIVE IMMUNOGLOBULIN HEAVY CHAIN VARIABLE REGION (FRAGMENT).  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=98277139; PubMed=9614934;  
 RA Wu X., Liu B., Van der Merwe P.L., Kalis N.N., Berney S.M.,  
 RT "Myosin-reactive autoantibodies in rheumatic carditis and normal  
 RT fetus.";

RL Clin. Immunol. Immunopathol. 87:184-192(1998).  
 DR EMBL; AF035023; AAD56259.1; -.  
 DR INTERPRO; IPR003006; -.  
 DR PFAM; PF00047; 1g; 1.  
 FT NON\_TER 1  
 FT NON\_TER 1  
 SQ SEQUENCE 118 AA; 12843 MW; D0633949F2AC149D CRC64;

Query Match 81.0%; Score 34; DB 4; Length 118;  
 Best Local Similarity 87.5%; Pred. No. 26;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GGLVOPG 8  
 || |||||  
 DB 8 GGLVOPG 15

## RESULT 8

O9UL72 PRELIMINARY; PRT; 118 AA.  
 AC O9UL72;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)  
 DE MYOSIN-REACTIVE IMMUNOGLOBULIN HEAVY CHAIN VARIABLE REGION (FRAGMENT).  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=98277139; PubMed=9614934;  
 RA Wu X., Liu B., Van der Merwe P.L., Kalis N.N., Berney S.M.,  
 RT "Myosin-reactive autoantibodies in rheumatic carditis and normal  
 RT fetus.";  
 RL Clin. Immunol. Immunopathol. 87:184-192(1998).  
 DR EMBL; AF035042; AAD56278.1; -.  
 DR HSSP; P01772; 2F84.  
 DR INTERPRO; IPR003006; -.  
 DR PFAM; PF00047; 1g; 1.  
 FT NON\_TER 1  
 FT 118  
 FT NON\_TER 1  
 SQ SEQUENCE 118 AA; 12872 MW; B4D1A5944B2D5CCA CRC64;

Query Match 81.0%; Score 34; DB 4; Length 118;  
 Best Local Similarity 87.5%; Pred. No. 26;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GGLVOPG 8  
 || |||||  
 DB 8 GGLVOPG 15

## RESULT 9

O9PJT5 PRELIMINARY; PRT; 238 AA.  
 AC O9PJT5;  
 DT 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)  
 DE HYPOTHETICAL PROTEIN TC0742.  
 GN TC0742.  
 OS Chlamydia muridarum.  
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
 OX NCBI\_TaxID=83560;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MOPN / NIGG;  
 RX MEDLINE=20150255; PubMed=10684935;  
 RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,

RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,  
 RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,  
 RA Gwin M., Nelson W., Deboy R., Kolonay J., McClarty G., Salzberg S.L.,  
 RA Eisen J., Fraser C.M.;  
 RT "Genome sequences of Chlamydia trachomatis MOpn and Chlamydia  
 RT pneumoniae AR39.";  
 RN Nucleic Acids Res. 28:1397-1406(2000).  
 RL [2]  
 RP SEQUENCE FROM N.A.  
 RA Read T.D., Brunham R., Shen C., Gill S.R., Heidelberg J.F., White O.,  
 RA Hickey E.K., Peterson J., Umayam L.A., Utterback T., Berry K.,  
 RA Bass S., Linher K., Weidman J., Khouri H., Craven B., Bowman C.,  
 RA Dodson R., Gwin M., Nelson W., Deboy R., Kolonay J., McClarty G.,  
 RA Salzberg S.L., Eisen J., Fraser C.M.;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AE002342; AAF39551.1; -;  
 DR TIGR: TC0742; -;  
 DR INTERPRO: IPR002876; -;  
 DR PFAM: PF01709; DUF28; 1.  
 DR Hypothetical protein.  
 KM  
 SQ SEQUENCE 238 AA; 26486 MW; C5D5657B0909264C CRC64;

Query Match 81.0%; Score 34; DB 2; Length 238;  
 Best Local Similarity 75.0%; Pred. No. 56;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 GGLVOPG 8  
 11:11:11  
 DB 120 GGLVOPG 127

RESULT 10  
 057998  
 ID 057998 PRELIMINARY; PRT; 257 AA.  
 AC 057998;  
 DT 01-AUG-1998 (TrEMBLrel. 07, Created)  
 DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)  
 DE HYPOTHEMETICAL 28.7 KDA PROTEIN PH0260.  
 GN PH0260.  
 OS Pyrococcus horikoshii.  
 OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.  
 OC NCBI\_TaxID=53953;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=OT3;  
 RX MEDLINE=98344137; PubMed=9679194;  
 RA Kawarabayashi Y., Sawada M., Horikawa H., Hatakeyama Y., Hino Y.,  
 RA Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,  
 RA Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y.,  
 RA Funahashi T., Tanaka T., Kudoh Y., Yamazaki J., Kushiida N., Oguchi A.,  
 RA Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,  
 RA Mesuchi Y., Shizuya H., Kikuchi H.;  
 RT "Complete sequence and gene organization of the genome of a hyper-  
 RT thermophilic archaeobacterium, Pyrococcus horikoshii OT3.";  
 RL DNA Res. 5:55-76(1998).  
 DR EMBL: AP000001; BAA29332.1; -;  
 DR INTERPRO: IPR001279; -;  
 DR PFAM: PF00753; lactamase\_B; 1.  
 KM Hypothetical protein.  
 SQ SEQUENCE 257 AA; 28680 MW; C9249670AB73FA7E CRC64;

Query Match 81.0%; Score 34; DB 1; Length 257;  
 Best Local Similarity 62.5%; Pred. No. 60;  
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 GGLVOPG 8  
 11:11:11  
 DB 44 GGLVOPG 51

RESULT 11  
 09V2H0  
 ID 09V2H0 PRELIMINARY; PRT; 272 AA.  
 AC 09V2H0;  
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
 DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)  
 DE HYPOTHEMETICAL 30.4 KDA PROTEIN.  
 GN PAB2280.  
 OS Pyrococcus abyssi.  
 OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.  
 OC NCBI\_TaxID=29292;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ORSAY;  
 RA Hellig R.;  
 RT "Pyrococcus abyssi genome sequence: Insights into archaeal chromosome  
 RT structure and evolution.";  
 RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AJ248283; CAB49028.1; -;  
 DR Hypothetical protein.  
 KM  
 SQ SEQUENCE 272 AA; 30372 MW; AE0250A0399E01AA CRC64;

Query Match 81.0%; Score 34; DB 1; Length 272;  
 Best Local Similarity 75.0%; Pred. No. 64;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 GGLVOPG 8  
 11:11:11  
 DB 179 GGLVOPG 186

RESULT 12  
 073907  
 ID 073907 PRELIMINARY; PRT; 323 AA.  
 AC 073907;  
 DT 01-AUG-1998 (TrEMBLrel. 07, Created)  
 DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)  
 DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)  
 DE CNKX-2.3 PROTEIN.  
 GN CNKX-2.3.  
 OS Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 OC Gallus.  
 OC NCBI\_TaxID=9031;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=HEART;  
 RX MEDLINE=96391182; PubMed=8798155;  
 RA Buchberger A., Padst O., Brandt T., Seidl K., Arnold H.H.;  
 RT "Chick CNKX-2.3 represents a novel family member of vertebrate  
 RT homologues to the Drosophila homeobox gene tinman: differential  
 RT expression of CNKX-2.3 and CNKX-2.5 during heart and gut  
 RT development.";  
 RL Mech. Dev. 56:151-163(1996).  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).  
 DR EMBL: X97667; CAA66257.1; -;  
 DR HSSP: P23441; LEFT.  
 DR INTERPRO: IPR001356; -;  
 DR PFAM: PF00046; homeobox; 1.  
 DR PRINTS: PR00024; HOMEBOX.  
 DR PROSITE: PS00027; HOMEBOX\_1; 1.  
 DR PROSITE: PS50071; HOMEBOX\_2; 1.  
 KM Homeobox; DNA-binding; Nuclear protein.  
 SQ SEQUENCE 323 AA; 36034 MW; 09B644D7D54470CB CRC64;

Query Match 81.0%; Score 34; DB 13; Length 323;  
 Best Local Similarity 75.0%; Pred. No. 78;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;



OY 1 GGLVQPG 8  
 111:111  
 DB 308 GGLVQPG 315

RESULT 13  
 ID 035126 PRELIMINARY; PRT: 1175 AA.  
 AC 035126.

DT 01-JAN-1998 (TREMBLrel. 05, Created)  
 DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)  
 DE DRPLA.

GN DRPLA.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;

RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Ansaari-Lari M.A., Oeltjen J.C., Schwartz S., Zhang Z., Muzny D.M.,  
 RA Lu J., Gorrell J.H., Chinnait A.C., Belmont J.W., Miller W.,  
 RA Gibbs R.A.;  
 RN Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.

RP [2]  
 RP SEQUENCE FROM N.A.  
 RA Muzny D., Ansaari-Lari M.A., Timms K.M., Yu W., Dugan S., Lu J.,  
 RA Shen Y., Rowland K., Liu W., Perez L., Ding Y., Haywood M., Jain A.,  
 RA Leal B., Logan O., Nguyen V., Savage L., Shen H., Worley K., Chen E.,  
 RA Forcum J., Aronsen A.D., Chiu M.W., Gorrell J.H., Brundage E., Di W.,  
 RA Chinnait C., Nelson D., Gibbs R.A.;  
 RN Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.

DR EMBL: AC002397; AAC36003.1; -.  
 DR MGD: MGI:104725; Drpla.  
 DR INTERPRO: IPR002951; -.  
 DR PRINTS: PR01222; ATROPHIN.  
 SO SEQUENCE 1175 AA; 123722 MW; 8BEFFAB75DDC0F36 CRC64;

Query Match 81.0%; Score 34; DB 11; Length 1175;  
 Best Local Similarity 75.0%; Pred. No. 3.2e+02;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 GGLVQPG 8  
 111:111  
 DB 960 GGLVQPG 967

RESULT 14  
 ID P70200 PRELIMINARY; PRT: 1175 AA.  
 AC P70200.

DT 01-FEB-1997 (TREMBLrel. 02, Created)  
 DT 01-FEB-1997 (TREMBLrel. 02, Last sequence update)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)  
 DE DRPLA PROTEIN.

GN DRPLA.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;

RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE-9724514; PubMed-9070948;  
 RA OYake M., Onodera O., Shirolshi T., Takano H., Takahashi Y.,  
 RA Komonami R., Moriwaki K., Ikeuchi T., Igarashi S., Tanaka H.,  
 RA Tsuji S.;  
 RT "Molecular cloning of murine homologue dentatorubral-pallidolysian  
 atrophy (DRPLA) cDNA: strong conservation of a polymorphic CAG repeat  
 in the murine gene."  
 RT Genomics 40:205-207(1997).  
 DR EMBL: D87744; BAA13450.1; -.  
 DR MGD: MGI:104725; Drpla.

DR INTERPRO: IPR002951; -.  
 DR PRINTS: PR01222; ATROPHIN.  
 SO SEQUENCE 1175 AA; 123641 MW; DEBDCCE0FC05CBF6 CRC64;

Query Match 81.0%; Score 34; DB 11; Length 1175;  
 Best Local Similarity 75.0%; Pred. No. 3.2e+02;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 GGLVQPG 8  
 111:111  
 DB 960 GGLVQPG 967

RESULT 15  
 ID 099621 PRELIMINARY; PRT: 1190 AA.  
 AC 099621.

DT 01-MAY-1997 (TREMBLrel. 03, Created)  
 DT 01-MAY-1997 (TREMBLrel. 03, Last sequence update)  
 DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)  
 DE DRPLA.

GN HUMDRPLA.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 OX NCBI\_TaxID=9606;

RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE-96303695; PubMed=8723724;  
 RA Ansaari-Lari M.A., Muzny D.M., Lu J., Lu F., Lilley C.E., Spanos S.,  
 RA Malley T., Gibbs R.A.;  
 RT "A gene-rich cluster between the CD4 and tritosephosphate isomerase  
 RT genes at human chromosome 12p13."  
 RL Genome Res. 6:314-326(1996).  
 RN [2]  
 RP SEQUENCE FROM N.A.

RX MEDLINE-97228904; PubMed=9074930;  
 RA Ansaari-Lari M.A., Shen Y., Muzny D.M., Lee W., Gibbs R.A.;  
 RT "Large-scale sequencing in human chromosome 12p13: experimental and  
 RT computational gene structure determination."  
 RL Genome Res. 7:268-280(1997).  
 DR EMBL: U47924; CAB35665.1; -.  
 DR INTERPRO: IPR002951; -.  
 DR PRINTS: PR01222; ATROPHIN.  
 SO SEQUENCE 1190 AA; 125413 MW; B47603486C672637 CRC64;

Query Match 81.0%; Score 34; DB 4; Length 1190;  
 Best Local Similarity 75.0%; Pred. No. 3.2e+02;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 GGLVQPG 8  
 111:111  
 DB 975 GGLVQPG 982

Search completed: June 13, 2001, 14:20:37  
 Job time: 737 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

## OM protein - protein search, using sw model

Run on: June 13, 2001, 14:16:37 ; Search time 118.55 Seconds  
(without alignments)  
1.296 Million cell updates/sec

Title: PCT-US01-05825a-19  
Perfect score: 42  
Sequence: 1 GGLVQPG 8

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 185757 seqs, 19210857 residues

Total number of hits satisfying chosen parameters: 185757

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep:\*  
2: /cgn2\_6/ptodata/2/1aa/5B.COMB.pep:\*  
3: /cgn2\_6/ptodata/2/1aa/6A.COMB.pep:\*  
4: /cgn2\_6/ptodata/2/1aa/6B.COMB.pep:\*  
5: /cgn2\_6/ptodata/2/1aa/PCITUS.COMB.pep:\*  
6: /cgn2\_6/ptodata/2/1aa/Backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	39	92.9	20	2	US-08-859-931A-4
2	37	88.1	118	3	US-08-545-809A-108
3	37	88.1	123	2	US-08-428-197-38
4	37	88.1	123	5	PCT-US93-10555-38
5	36	85.7	118	3	US-08-545-809A-125
6	36	85.7	489	1	US-08-434-702-4
7	35	83.3	119	2	US-08-475-000-16
8	35	83.3	119	2	US-08-483-199-16
9	35	83.3	119	2	US-08-484-508-16
10	34	81.0	15	1	US-08-331-398A-66
11	34	81.0	15	2	US-08-632-358-42
12	34	81.0	15	2	US-08-331-397B-66
13	34	81.0	15	2	US-08-759-804A-65
14	34	81.0	18	1	US-08-331-398A-55
15	34	81.0	18	1	US-08-401-908-14
16	34	81.0	18	2	US-08-331-397B-55
17	34	81.0	18	2	US-08-759-804A-54
18	34	81.0	20	2	US-08-859-931A-2
19	34	81.0	26	1	US-08-053-451B-113
20	34	81.0	26	1	US-08-471-780C-80
21	34	81.0	26	1	US-08-467-282B-80
22	34	81.0	26	2	US-08-471-282A-80
23	34	81.0	26	2	US-08-466-710C-80
24	34	81.0	26	3	US-08-468-739C-80
25	34	81.0	29	1	US-08-471-780C-80
26	34	81.0	29	1	US-08-467-282B-119
27	34	81.0	29	2	US-08-471-282A-119

28	34	81.0	29	2	US-08-466-710C-119	Sequence 119, App
29	34	81.0	29	3	US-08-468-739C-119	Sequence 119, App
30	34	81.0	30	1	US-07-988-925-7	Sequence 7, Appl1
31	34	81.0	30	1	US-07-977-696C-75	Sequence 75, Appl1
32	34	81.0	30	1	US-08-129-930B-75	Sequence 75, Appl1
33	34	81.0	30	2	US-08-362-780-7	Sequence 92, Appl1
34	34	81.0	30	2	US-08-765-783A-92	Sequence 17, Appl1
35	34	81.0	30	2	US-08-470-139-17	Sequence 36, Appl1
36	34	81.0	67	1	US-08-162-102C-36	Sequence 126, App
37	34	81.0	86	2	US-08-053-451B-126	Sequence 21, Appl1
38	34	81.0	87	1	US-08-497-312-21	Sequence 105, App
39	34	81.0	87	2	US-08-765-783A-105	Sequence 116, App
40	34	81.0	98	1	US-08-211-202-116	Sequence 37, Appl1
41	34	81.0	98	1	US-07-942-245-37	Sequence 48, Appl1
42	34	81.0	98	2	US-08-428-197-48	Sequence 48, Appl1
43	34	81.0	98	2	US-08-665-202-31	Sequence 48, Appl1
44	34	81.0	98	5	PCT-US93-10555-48	Sequence 4, Appl1
45	34	81.0	108	2	US-08-428-197-4	

## ALIGNMENTS

RESULT 1  
US-08-859-931A-4  
; Sequence 4, Application US/08859931A  
; Patent No. 5945510  
; GENERAL INFORMATION:  
; APPLICANT: FASANO, Alessio  
; TITLE OF INVENTION: SUBSTANTIALLY PURE ZONULIN, A  
; TITLE OF INVENTION: PHYSIOLOGICAL MODULATOR OF  
; TITLE OF INVENTION: MAMMALIAN TIGHT JUNCTIONS  
; NUMBER OF SEQUENCES: 9  
; CORRESPONDENCE ADDRESSES:  
; ADDRESS: SUGHRUE, MIOM, ZINN, MACPEAK & SEAS  
; STREET: 2100 Pennsylvania Avenue, N.W., Suite 800  
; CITY: Washington, D.C.  
; STATE: D.C.  
; COUNTRY: U.S.A.  
; ZIP: 20037  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/859,931A  
; FILING DATE: 21 MAY 1997  
; CLASSIFICATION: 514  
; ATTORNEY/AGENT INFORMATION:  
; NAME: KIT, Gordon  
; REGISTRATION NUMBER: 30,764  
; REFERENCE/DOCKET NUMBER: A-6901  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 293-7060  
; TELEFAX: (202) 293-7860  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 20 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; HYPOTHEICAL: NO  
; US-08-859-931A-4

Query Match 92.9%; Score 39; DB 2; Length 20;  
Best Local Similarity 87.5%; Pred. No. 0.28;  
Matches 7; Conservative 1; Mismatches 0; Gaps 0;  
DB 8 GGLVQPG 15

RESULT 2  
US-08-545-809A-108  
Sequence 108, Application US/08545809A  
Patent No. 6096878  
GENERAL INFORMATION:  
APPLICANT: Honjo, Tasuku  
TITLE OF INVENTION: HUMAN IMMUNOGLOBULIN VH GENE  
NUMBER OF SEQUENCES: 145  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson, P.C.  
STREET: 225 Franklin Street  
CITY: Boston  
STATE: MA  
COUNTRY: US  
ZIP: 02110-2804  
COMPUTER READABLE FORM:  
MEDIUM TYPE: diskette  
COMPUTER: IBM compatible  
OPERATING SYSTEM: Windows95  
SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/545,809A  
FILING DATE: 27-MAR-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/JP93/00603  
FILING DATE: 10-MAY-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Freeman, John W.  
REGISTRATION NUMBER: 29,066  
REFERENCE/DOCKET NUMBER: 06501/004001  
TELEPHONE: 617-542-5070  
TELEFAX: 617-542-8906  
TELEX: 200154  
INFORMATION FOR SEQ ID NO: 108:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 118 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-545-809A-108

Query Match 88.1%; Score 37; DB 3; Length 118;  
Best Local Similarity 87.5%; Pred. NO. 4;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGLVQPG 8  
111111  
DB 27 GGLVQPG 34

RESULT 3  
US-08-428-197-38  
Sequence 38, Application US/08428197  
Patent No. 5891438  
GENERAL INFORMATION:  
APPLICANT: SILVERMAN, GREGG J.  
TITLE OF INVENTION: METHOD FOR STIMULATING PRODUCTION OF  
TITLE OF INVENTION: VARIABLE REGION GENE FAMILY RESTRICTED ANTIBODIES THROUGH  
TITLE OF INVENTION: VACCINATION WITH A B-CELL SUPERANTIGEN AND CONJUGATES  
TITLE OF INVENTION: THEREOF  
NUMBER OF SEQUENCES: 51  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Spensley Horn Juba & Lubitz  
STREET: 1880 Century Park East - Suite 500  
CITY: Los Angeles  
STATE: California  
COUNTRY: USA

ZIP: 90067  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/428,197  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US93/10555  
FILING DATE: 29-OCT-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Howells, Stacy L.  
REGISTRATION NUMBER: 34,842  
REFERENCE/DOCKET NUMBER: FD-2630  
TELEPHONE: (619) 455-5110  
TELEFAX: (619) 455-5110  
INFORMATION FOR SEQ ID NO: 38:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 123 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-428-197-38

Query Match 88.1%; Score 37; DB 2; Length 123;  
Best Local Similarity 87.5%; Pred. NO. 4.2;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGLVQPG 8  
111111  
DB 8 GGLVQPG 15

RESULT 4  
PCT-US93-10555-38  
Sequence 38, Application PC/TUS9310555  
GENERAL INFORMATION:  
APPLICANT: SILVERMAN, GREGG J.  
TITLE OF INVENTION: METHOD FOR STIMULATING PRODUCTION OF  
TITLE OF INVENTION: VARIABLE REGION GENE FAMILY RESTRICTED ANTIBODIES THROUGH  
TITLE OF INVENTION: VACCINATION WITH A B-CELL SUPERANTIGEN AND CONJUGATES  
TITLE OF INVENTION: THEREOF  
NUMBER OF SEQUENCES: 51  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Spensley Horn Juba & Lubitz  
STREET: 1880 Century Park East - Suite 500  
CITY: Los Angeles  
STATE: California  
COUNTRY: USA  
ZIP: 90067  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US93/10555  
FILING DATE: 29-OCT-1993  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Howells, Stacy L.  
REGISTRATION NUMBER: 34,842  
REFERENCE/DOCKET NUMBER: FD-2630  
TELEPHONE: (619) 455-5110  
TELEFAX: (619) 455-5110  
INFORMATION FOR SEQ ID NO: 38:  
SEQUENCE CHARACTERISTICS:

LENGTH: 123 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
PCT-US93-10555-38

Query Match 88.1%; Score 37; DB 5; Length 123;  
Best Local Similarity 87.5%; Pred. No. 4.2;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GGLVOPG 8  
||:||||  
Db 8 GGLVOPG 15

RESULT 5  
US-08-545-809A-125  
; Sequence 125, Application US/08545809A  
; Patent No. 6096878  
; GENERAL INFORMATION:  
; APPLICANT: Honjo, Tasuku  
; APPLICANT: Matsuda, Fumihiko  
; TITLE OF INVENTION: HUMAN IMMUNOGLOBULIN VH GENE  
; TITLE OF INVENTION: SEGMENTS AND DNA FRAGMENTS CONTAINING THE SAME  
; NUMBER OF SEQUENCES: 145  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson, P.C.  
; STREET: 225 Franklin Street  
; CITY: Boston  
; STATE: MA  
; COUNTRY: US  
; ZIP: 02110-2804  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; OPERATING SYSTEM: Windows95  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/545,809A  
; FILING DATE: 27-MAR-1996  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/JP93/00603  
; FILING DATE: 10-MAY-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Freeman, John W.  
; REGISTRATION NUMBER: 29,066  
; REFERENCE/DOCKET NUMBER: 06501/004001  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617-542-5070  
; TELEFAX: 617-542-8906  
; TELEX: 200134  
; INFORMATION FOR SEQ ID NO: 125:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 118 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-545-809A-125

Query Match 85.7%; Score 36; DB 3; Length 118;  
Best Local Similarity 75.0%; Pred. No. 6.1;  
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 GGLVOPG 8  
||:||||  
Db 27 GGLVOPG 34

RESULT 6  
US-08-434-702-4  
; Sequence 4, Application US/08434702

; Patent No. 5554743  
; GENERAL INFORMATION:  
; APPLICANT: Bennett, Alan B.  
; APPLICANT: Fischer, Robert L.  
; APPLICANT: Lashbrook, Coralie  
; APPLICANT: Giovannini, James  
; TITLE OF INVENTION: Endo-1,4-beta-Glucanase Genes and Their  
; TITLE OF INVENTION: Use in Plants  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Townsend and Townsend and Crew  
; STREET: One Market Plaza, Stewart Street Tower  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94105-1492  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/434,702  
; FILING DATE: 04-MAY-1995  
; CLASSIFICATION: 536  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/271,883  
; FILING DATE: 07-JUL-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/687,466  
; FILING DATE: 18-APR-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/511,417  
; FILING DATE: 20-APR-1990  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Bastian, Kevin L.  
; REGISTRATION NUMBER: 34,774  
; REFERENCE/DOCKET NUMBER: 02307E-304300S  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 543-9600  
; TELEFAX: (415) 543-5043  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 489 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-434-702-4

Query Match 85.7%; Score 36; DB 1; Length 489;  
Best Local Similarity 75.0%; Pred. No. 26;  
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 GGLVOPG 8  
||:||||  
Db 327 GGLVOPG 334

RESULT 7  
US-08-475-000-16  
; Sequence 16, Application US/08475000  
; Patent No. 5811267  
; GENERAL INFORMATION:  
; APPLICANT: RING, DAVID B.  
; TITLE OF INVENTION: ANTIGEN-BINDING SITES OF ANTIBODY  
; TITLE OF INVENTION: MOLECULES SPECIFIC FOR CANCER ANTIGENS  
; NUMBER OF SEQUENCES: 18  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: CHIRON CORPORATION  
; STREET: INTELLECTUAL PROPERTY R440 - P.O. BOX 8097  
; CITY: Emeryville  
; STATE: CA

COUNTRY: USA  
ZIP: 94662-8097  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/475,000  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: SAVERIDE, PAUL  
REGISTRATION NUMBER: 36,914  
REFERENCE/DOCKET NUMBER: 0850.007  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (510) 601-2585  
TELEFAX: (510) 655-3542  
INFORMATION FOR SEQ ID NO: 16:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 119 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-475-000-16

Query Match 83.3%; Score 35; DB 2; Length 119;  
Best Local Similarity 75.0%; Pred. No. 9.4;  
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGLVOPG 8  
||:|:|  
DB 8 GGVLRPG 15

RESULT 8  
US-08-483-199-16  
Sequence 16, Application US/08483199  
Patent No. 5849877  
GENERAL INFORMATION:  
APPLICANT: RING, DAVID B.  
TITLE OF INVENTION: ANTIGEN-BINDING SITES OF ANTIBODY  
TITLE OF INVENTION: MOLECULES SPECIFIC FOR CANCER ANTIGENS  
NUMBER OF SEQUENCES: 18  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CHIRON CORPORATION  
STREET: INTELLECTUAL PROPERTY - R440, P.O. BOX 8097  
CITY: Emeryville  
STATE: CA  
COUNTRY: USA  
ZIP: 94662-8097  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/483,199  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: SAVERIDE, PAUL  
REGISTRATION NUMBER: 36,914  
REFERENCE/DOCKET NUMBER: 0850.009  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (510) 601-2585  
TELEFAX: (510) 655-3542  
INFORMATION FOR SEQ ID NO: 16:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 119 amino acids  
TYPE: amino acid  
TOPOLOGY: linear

MOLECULE TYPE: protein  
US-08-483-199-16

Query Match 83.3%; Score 35; DB 2; Length 119;  
Best Local Similarity 75.0%; Pred. No. 9.4;  
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGLVOPG 8  
||:|:|  
DB 8 GGVLRPG 15

RESULT 9  
US-08-484-508-16  
Sequence 16, Application US/08484508  
Patent No. 5948647  
GENERAL INFORMATION:  
APPLICANT: RING, DAVID B.  
TITLE OF INVENTION: ANTIGEN-BINDING SITES OF ANTIBODY  
TITLE OF INVENTION: MOLECULES SPECIFIC FOR CANCER ANTIGENS  
NUMBER OF SEQUENCES: 18  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CHIRON CORPORATION  
STREET: INTELLECTUAL PROPERTY - R440, P.O. BOX 8097  
CITY: Emeryville  
STATE: CA  
COUNTRY: USA  
ZIP: 94662-8097  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/484,508  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:  
NAME: SAVERIDE, PAUL  
REGISTRATION NUMBER: 36,914  
REFERENCE/DOCKET NUMBER: 0850.008  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (510) 601-585  
TELEFAX: (510) 655-3542  
INFORMATION FOR SEQ ID NO: 16:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 119 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-484-508-16

Query Match 83.3%; Score 35; DB 2; Length 119;  
Best Local Similarity 75.0%; Pred. No. 9.4;  
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGLVOPG 8  
||:|:|  
DB 8 GGVLRPG 15

RESULT 10  
US-08-331-398A-66  
Sequence 66, Application US/08331398A  
Patent No. 5608039  
GENERAL INFORMATION:  
APPLICANT: Pastan, Ira  
APPLICANT: Willingham, Mark  
APPLICANT: Fitzgerald, David  
APPLICANT: Brinkmann, Ulrich  
APPLICANT: Pal, Lee

TITLE OF INVENTION: Single Chain B3 Antibody Fusion Proteins  
TITLE OF INVENTION: and Their Uses (as amended)  
NUMBER OF SEQUENCES: 68  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend and Crew  
STREET: One Market Plaza, Stewart Street Plaza  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94105-1492  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/331,398A  
FILING DATE: 28-OCT-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/767,331  
FILING DATE: 30-SEP-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/596,289  
FILING DATE: 12-OCT-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: Hunter, Tom  
REGISTRATION NUMBER: 38,498  
REFERENCE/DOCKET NUMBER: 015280-126110US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 543-9600  
TELEFAX: (415) 543-5043  
INFORMATION FOR SEQ ID NO: 66:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 15 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-331-398A-66

Query Match 81.0%; Score 34; DB 1; Length 15;  
Best local Similarity 87.5%; Pred. No. 1.7;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GGLVOPG 8  
Db 8 GGLVOPG 15

RESULT 11  
US-08-652-558-42  
Sequence 42, Application US/08652558  
Patent No. 5861155  
GENERAL INFORMATION:  
APPLICANT: LIN, AUGUSTINE YEE-THARN  
TITLE OF INVENTION: HUMANIZED ANTIBODIES AND USES  
TITLE OF INVENTION: THEREOF  
NUMBER OF SEQUENCES: 52  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BANNER & WITCOFF  
STREET: 75 STATE STREET, 23RD FLOOR  
CITY: BOSTON  
STATE: MASSACHUSETTS  
COUNTRY: USA  
ZIP: 02109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 inch diskette, 1.4 Mb storage  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: WordPerfect 6.1  
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/652,558  
FILING DATE: JUNE 6, 1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/IB94/00387  
FILING DATE: NOVEMBER 21, 1994  
ATTORNEY/AGENT INFORMATION:  
NAME: YANKWICH, LEON R.  
REGISTRATION NUMBER: 30,237  
REFERENCE/DOCKET NUMBER: 95,497-L  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-345-9100  
TELEFAX: 617-345-9111  
INFORMATION FOR SEQ ID NO: 42:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 15 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-652-558-42

Query Match 81.0%; Score 34; DB 2; Length 15;  
Best local Similarity 87.5%; Pred. No. 1.7;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GGLVOPG 8  
Db 8 GGLVOPG 15

RESULT 12  
US-08-331-397B-66  
Sequence 66, Application US/08331397B  
Patent No. 5981726  
GENERAL INFORMATION:  
APPLICANT: Pastan, Ira  
TITLE OF INVENTION: Chimeric and Mutationally Stabilized Tumor-  
TITLE OF INVENTION: Specific Antibody Fragments, Fusion Proteins, and Uses  
TITLE OF INVENTION: Thereof  
NUMBER OF SEQUENCES: 68  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend and Crew  
STREET: One Market Plaza, Stewart Street Plaza  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94105-1492  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/331,397B  
FILING DATE: 28-OCT-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/767,331  
FILING DATE: 30-SEP-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/596,289  
FILING DATE: 12-OCT-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: Hunter, Tom  
REGISTRATION NUMBER: 38,498  
REFERENCE/DOCKET NUMBER: 015280-126120US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 543-9600  
TELEFAX: (415) 543-5043  
INFORMATION FOR SEQ ID NO: 66:  
SEQUENCE CHARACTERISTICS:

LENGTH: 15 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-331-397B-66

Query Match 81.0%; Score 34; DB 2; Length 15;  
Best Local Similarity 87.5%; Pred. No. 1.7;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGLVOPG 8  
1111111  
Db 8 GGLVOPG 15

RESULT 13  
US-08-759-804A-65  
Sequence 65, Application US/08759804A  
Patent No. 5990296

GENERAL INFORMATION:

APPLICANT: Pastan, Ira  
APPLICANT: Willingham, Mark  
APPLICANT: Fitzgerald, David J.  
APPLICANT: Brinkmann, Ulrich  
APPLICANT: Pal, Lee  
TITLE OF INVENTION: Tumor-Specific Antibody Fragments,  
TITLE OF INVENTION: Fusion Proteins, and Uses Thereof  
NUMBER OF SEQUENCES: 68  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Center, Eighth Floor  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-3834

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/759,804A  
FILING DATE: 03-DEC-1996  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/331,398  
FILING DATE: 28-OCT-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/767,331  
FILING DATE: 30-SEP-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/596,289  
FILING DATE: 12-OCT-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: Weber, Ellen L.  
REGISTRATION NUMBER: 32,762  
REFERENCE/DOCKET NUMBER: 015280-126140US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 576-0200  
TELEFAX: (415) 576-0300  
INFORMATION FOR SEQ ID NO: 65:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 15 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-759-804A-65

Query Match 81.0%; Score 34; DB 2; Length 15;

Best Local Similarity 87.5%; Pred. No. 1.7;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGLVOPG 8  
1111111  
Db 8 GGLVOPG 15

RESULT 14  
US-08-331-398A-55  
Sequence 55, Application US/08331398A  
Patent No. 5608039

GENERAL INFORMATION:

APPLICANT: Pastan, Ira  
APPLICANT: Willingham, Mark  
APPLICANT: Fitzgerald, David  
APPLICANT: Brinkmann, Ulrich  
APPLICANT: Pal, Lee  
TITLE OF INVENTION: Single Chain B3 Antibody Fusion Proteins  
TITLE OF INVENTION: and Their Uses (as amended)  
NUMBER OF SEQUENCES: 68  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend and Crew  
STREET: One Market Plaza, Stewart Street Plaza  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94105-1492

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/331,398A  
FILING DATE: 28-OCT-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/767,331  
FILING DATE: 30-SEP-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/596,289  
FILING DATE: 12-OCT-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: Hunter, Tom  
REGISTRATION NUMBER: 38,498  
REFERENCE/DOCKET NUMBER: 015280-126110US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 543-9600  
TELEFAX: (415) 543-5043  
INFORMATION FOR SEQ ID NO: 55:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 18 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-331-398A-55

Query Match 81.0%; Score 34; DB 1; Length 18;  
Best Local Similarity 87.5%; Pred. No. 2.1;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGLVOPG 8  
1111111  
Db 8 GGLVOPG 15

RESULT 15  
US-08-401-908-14  
Sequence 14, Application US/08401908  
Patent No. 5684146



GENERAL INFORMATION:  
APPLICANT: Yoshinobu OKUNO et al.  
TITLE OF INVENTION: A DNA WHICH CODES FOR THE VARIABLE REGION OF  
TITLE OF INVENTION: ANTIBODY  
NUMBER OF SEQUENCES: 36  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Wenderoth, Lind & Ponack  
STREET: 805 Fifteenth Street, N.W., #700  
CITY: Washington  
STATE: D.C.  
COUNTRY: U.S.A.  
ZIP: 20005  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: Wordperfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/401,908  
FILING DATE: March 10, 1995  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Warren M. Cheek, Jr.  
REGISTRATION NUMBER: 33,367  
REFERENCE/DOCKET NUMBER:  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-371-8850  
TELEFAX:  
TELEX:  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 18  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
ORIGINAL SOURCE: heavy chain of C179  
US-08-401-908-14

Query Match 81.0%; Score 34; DB 1; Length 18;  
Best Local Similarity 87.5%; Pred. No. 2.1;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 GGLVQPG 8  
|||  
Db 8 GGGLVQPG 15

Search completed: June 13, 2001, 14:16:37  
Job time: 498 sec

*This Page Blank (uspto)*